

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 29, 2002, 08:20:28 ; Search time 1832.12 Seconds

(without alignments)
11719.007 Million cell updates/sec

Title: US-09-664-186-4

Perfect score: 1026

Sequence: 1 gttgaagaagaagaaacatt.....ccctcaaggagcctctctga 1026

Scoring table: IDENTITY_NUC

Searched: 1797656 seqs, 10463268293 residues

al number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

GenEmbl: *

1: gb_ba: *
2: gb_hgt: *
3: gb_in: *
4: gb_om: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_to: *
11: gb_sts: *
12: gb_sy: *
13: gb_un: *
14: gb_vl: *
15: em_ba: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_mu: *
20: em_om: *
21: em_or: *
22: em_ov: *
23: em_pat: *
24: em_ph: *
25: em_pl: *
26: em_ro: *
27: em_sts: *
28: em_un: *
29: em_vl: *
30: em_hgt_hum: *
31: em_hgt_inv: *
32: em_hgt_other: *
33: em_hgtg_inv: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result Query Score Match Length DB ID Description

1	1026	100.0	1026	6	AR139424	AR139424 Sequence
2	1026	100.0	1600	1	U89376	U89376 Thermus sp.
3	1026	100.0	5849	1	AF013571	AF013571 Thermus s
4	1026	100.0	5849	6	AR139426	AR139426 Sequence
5	59.8	5.8	38962	1	SC2E1	AL023797 Streptomy
6	56.8	5.5	13347	1	SCC42	AL135692 Streptomy
7	54.6	5.3	35576	1	SCD46	AL1391406 Streptomy
8	53.6	5.2	41622	1	SCD25	AL118514 Streptomy
9	53.2	5.2	100772	2	AC105260	AC105260 Oryza sat
10	53.2	5.2	167049	2	AP004746	AP004746 Oryza sat
11	52.6	5.1	7459	1	SCB82	AL445327 Streptomy
12	52.6	5.1	34983	1	SCM11	AL133278 Streptomy
13	52.4	5.1	43147	1	SC4A10	AL109663 Streptomy
14	52.4	5.1	113193	1	AF357202	AF357202 Streptomy
15	52.2	5.1	188050	1	AL646072	AL646072 Ralstonia
16	51.8	5.0	19372	1	AE004443	AE004443 Pseudomon
17	51.6	5.0	1815	6	AX148314	AX148314 Sequence
18	51.6	5.0	2498	8	ZM095953	U95953 Zea mays vl
19	51.2	5.0	615	6	E04306	E04306 DNA encodin
20	51.2	5.0	915	1	TRHSUPDIS	D13387 Thermus aqu
21	51.2	5.0	6955	1	TRHIVB	D84646 Thermus aqu
22	50.8	4.9	23458	1	SCB22	AL355832 Streptomy
23	50.6	4.9	30941	1	SC4C6	AL079335 Streptomy
24	50.6	4.9	33810	1	SCB6	AL353832 Streptomy
25	50.4	4.9	10790	1	AE001973	AE001973 Deinococc
26	50.4	4.9	21220	1	SC1E6	AL033505 Streptomy
27	50.4	4.9	125020	9	AF429315	AF429315 Homo sapi
28	50.2	4.9	24397	1	AE004673	AE004673 Pseudomon
29	50.2	4.9	41906	1	SC5C7	AL031515 Streptomy
30	50.2	4.9	123980	2	AC103551	AC103551 Oryza sat
31	50	4.9	11187	1	AE001908	AE001908 Deinococc
32	50	4.9	39314	1	SGR300302	AJ300302 Streptomy
33	49.8	4.9	173356	2	AP004047	AP004047 Oryza sat
34	49.6	4.8	11808	1	SC9P2	AL035559 Streptomy
35	49.6	4.8	13814	1	AE007092	AE007092 Mycobacte
36	49.6	4.8	63033	1	MTV008	AL021246 Mycobacte
37	49.2	4.8	11790	1	SCC80	AL442143 Streptomy
38	49.2	4.8	13571	1	AE004982	AE004982 Halobacte
39	49.2	4.8	40104	1	SCP11	AL132662 Streptomy
40	49	4.8	1229	1	TC1DH	X04519 Thermus cal
41	49	4.8	12277	1	AF340166	AF340166 Streptomy
42	49	4.8	33320	1	SC9B10	AL009204 Streptomy
43	49	4.8	213050	1	AL646067	AL646067 Ralstonia
44	48.6	4.7	154746	14	HSV2HG52	Z86099 Herpes simp
45	48.4	4.7	2184	1	AF170451	AF170451 Rhizobium

ALIGNMENTS

RESULT 1
LOCUS AR139424 1026 bp DNA
DEFINITION Sequence 4 from patent US 6207377.
ACCESSION AR139424
VERSION AR139424.1 GI:14481920
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1026)
AUTHORS Wayne,J. and Xu,S.
TITLE Method for construction of thermus-E. coli shuttle vectors and identification of two Thermus plasmid replication origins
JOURNAL Patent: US 6207377-A 4 27-MAR-2001;
FEATURES
source Location/Qualifiers
BASE COUNT 221 a 323 c 285 g 197 t
ORIGIN

Query Match 100.0%; Score 1026; DB 6; Length 1026;

Best Local Similarity 100.0%; Pred. No. 3.5e-174;
Matches 1026; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 1 gtgaagaacgnaaaacattctttagaagaagctttagagagcttagaggaaccacagac 60
Db 1 gtgaagaacgnaaaacattctttagaagaagctttagagagcttagaggaaccacac 60
OY 61 aacacgaatgcacacgaagggtatagatagagggtacagaggaattctttagccacgac 120
Db 61 aacacgaatgcacacgaagggtatagatagagggtacagaggaattctttagccacgac 120
OY 121 cccctccagatgtagaggtgacgaataatcgctcgcaagggtttacataaccacaaagag 180
Db 121 cccctccagatgtagaggtgacgaataatcgctcgcaagggtttacataaccacaaagag 180
OY 181 gcaacttagatgtcttaccggaagaacacatagaggttcccttctctgttggggcc 240
Db 181 gcaacttagatgtcttaccggaagaacacatagaggttcccttctctgttggggcc 240
OY 241 cccctccacacacacagctgaaccccccgcttggaatgtaatgtaacagcgctccagagag 300
Db 241 cccctccacacacacagctgaaccccccgcttggaatgtaatgtaacagcgctccagagag 300
OY 301 cctcagaagctctggaactcctccagagaattgcgcgctccacacatccctcctacggcaac 360
Db 301 cctcagaagctctggaactcctccagagaattgcgcgctccacacatccctcctacggcaac 360
OY 361 cggagagctctggaagaagtgtagagcgctgcttcatatgctccctctgagatgttgcc 420
Db 361 cggagagctctggaagaagtgtagagcgctgcttcatatgctccctctgagatgttgcc 420
OY 421 ctcaaccttgagggtacacacgagacgcttcaacgcttggaagaaggtctctgagaanaag 480
Db 421 ctcaaccttgagggtacacacgagacgcttcaacgcttggaagaaggtctctgagaanaag 480
OY 481 ggcctggtggtgacacgagacgttcttcaacacacgttcaacggtgagcgcgagcgatcggc 540
Db 481 ggcctggtggtgacacgagacgttcttcaacacacgttcaacggtgagcgcgagcgatcggc 540
OY 541 aaccttggagcgtcagcgttagagccagaggaagccagacgtcacctctgagagacttacc 600
Db 541 aaccttggagcgtcagcgttagagccagaggaagccagacgtcacctctgagagacttacc 600
OY 601 taaccttgaggaacactgcacctagacaatggcaacgagcgctgctctccttcaactgggtc 660
Db 601 taaccttgaggaacactgcacctagacaatggcaacgagcgctgctctccttcaactgggtc 660
OY 661 aagccttaccaggaacacaggaataccgccccaccttgagacgtgctggtccttgggtcag 720
Db 661 aagccttaccaggaacacaggaataccgccccaccttgagacgtgctggtccttgggtcag 720
OY 721 ggaagaaagggtatgacccaacaacagacgctgagcgttgaacctggagcctatcctgtgc 780
Db 721 ggaagaaagggtatgacccaacaacagacgctgagcgttgaacctggagcctatcctgtgc 780
OY 781 ctccccagagtgtgagcgcttccaaactcccgcccttaccctcattgtacgtacatt 840
Db 781 ctccccagagtgtgagcgcttccaaactcccgcccttaccctcattgtacgtacatt 840
OY 841 ggcgattcctcctagatgacacgcttcaagaagcttctatgtagagcttgctgtggcgtg 900
Db 841 ggcgattcctcctagatgacacgcttcaagaagcttctatgtagagcttgctgtggcgtg 900
OY 901 gccaggggtgaaactcccgccgcaatattctatttgcctgcttcaatgacgggtattatccgagat 960
Db 901 gccaggggtgaaactcccgccgcaatattctatttgcctgcttcaatgacgggtattatccgagat 960
OY 961 taacaggaatggaactctgaacacgacggagagctacctaagtgaagaaccttaagaagggcc 1020
Db 961 taacaggaatggaactctgaacacgacggagagctacctaagtgaagaaccttaagaagggcc 1020
OY 1021 ttctgtga 1026
Db 1021 ttctgtga 1026

```

Db 1021 TCCTGA 1026

```

RESULT 2
LOCUS U89376 1600 bp DNA linear BCT 06-OCT-1997
DEFINITION Thermus sp. Y545 replication protein (rept) gene, complete cds.
ACCESSION U89376
VERSION U89376.1 GI:2465516
KEYWORDS
SOURCE
  Thermus sp. Y545.
  Bacteria; Thermus/Deinococcus group; Thermus group; Thermus.
REFERENCE
  1 (bases 1 to 1600)
  Wayne, J. and Xu, S.-Y.
  Identification of a thermophilic plasmid origin and its cloning
  within a new Thermus-E. coli shuttle vector
  Gene 195 (2), 321-328 (1997)
JOURNAL MEDLINE 97449309
REFERENCE 2 (bases 1 to 1600)
AUTHORS Wayne, J. and Xu, S.-Y.
TITLE Direct Submission
JOURNAL Submitted (12-FEB-1997) Research, New England Biolabs, 32 Tozer
          Road, Beverly, MA 01915, USA
FEATURES
  source
    1..1600
    /organism="Thermus sp. Y545"
    /strain="Y545"
    /db_xref="taxon:64227"
    220..225
    /note="putative"
    243..248
    /note="putative"
    278..283
    /note="putative"
    302..307
    /note="putative"
    332..336
    /gene="rept"
    332..1369
    /gene="rept"
    344..1369
    /gene="rept"
    /function="required for plasmid replication"
    /note="Rept"
    /codon_start=1
    /transl_table=11
    /product="replication protein"
    /protein_id="AB72144.1"
    /db_xref="GI:2465517"
    /translation="MKNEKTFEELYEALETHDNTDATRGSDGSEDFPIATDPPD
    GGAENLAGFTYOKALRIALPDENHEAFISVGA PTPAPAPPANVCOVQDPO
    KLEELIQLAKSTIPIGNRELRKVTYVFWPLEKALINTGRTOTVAMKVKLEK
    GLVATVLAHOTVNGERKALIGTLMAVRKRGKARLTLDYIYVPRNALDMANGVISEN
    WKAAYDGHGIRPLDVLVMAOGRVMPKTVAVDLGLVLPVEVERKLPALITLI
    ATYADLDLDRSRFRYAGLWAVARGELPAQYLFAVLMRVIRDYDGLTRGAVLY
    KTLKEAS"
    1158..1166
    /gene="rept"
    /note="putative; encodes Dnaa box"
    1293..1301
    /gene="rept"
    /note="putative; encodes Dnaa box"

```

BASE COUNT 340 a 487 c 458 g 315 t

Query Match 100.0%; Score 1026; DB 1; Length 1600;
Best Local Similarity 100.0%; Pred. No. 3.2e-174;
Matches 1026; Conservative 0; Mismatches 0; Indels 0; Gaps 0;


```

DB 4097 AACACGATGCCACTAGAGGGGGTACAGTAGGGGTGAGAGACTCTTCTGGCCACGAC 4038
|
|
|
OY 121 cccctccagatgaggtgcggaatacgcctcgagaggtctacataccaataaag 180
|
|
|
DB 4037 CCCCCCAGATGAGAGTCCGAAATCCCTCGCGAAGGGCTTACATACCAAAAAAGAG 3978
|
|
|
OY 181 gcaactagatgtcttaccgcggaagacctgagcttctctctctgttgggccc 240
|
|
|
DB 3977 GCACCTTAGATTGCTTTACCCGAGAAAGACATGAGGCTTTCTTCTCTGTGGGGCC 3918
|
|
|
OY 241 ccccccatacccaacgtgaacccccgttgggaagtatgtcaacgcgtccagagacgg 300
|
|
|
DB 3917 CCCCCATACACACACTGTAACCCCCCGTGGGAATGTAATGTAACCCGTCAGAGACGG 3858
|
|
|
OY 301 cctcagaagcttctggaactcctccagagatgtccgcgtccacacatccctcaggaac 360
|
|
|
DB 3857 CCTCAGAACCTTCTGGAAGCTCTCCAGAGATTGCCCGCTCCACATCCCTACGGCCAAC 3798
|
|
|
OY 361 cggagagctctgagagaggttgggagagctgtcttcatgtccctccctgagagatgtggcc 420
|
|
|
DB 3797 CGGAGACTCTGAGAGAGGTGGGAGCGGTGCTTCAATGCTCCCTCGAGATGTTGGCC 3738
|
|
|
OY 421 ctcaacctgagggtlcaaccgcgcagaccgttccacgccttggagaaggtlcttggaaaaag 480
|
|
|
DB 3737 CTCACACTGGGGGTCAACCCGGCAGACCGTCACGCTGGAAAGAGTCTTGAAGAAAAG 3678
|
|
|
OY 481 ggcctgtgtgcaacgcagctctcttccacaaacgcgttaaggggagagcgccgacatcgcc 540
|
|
|
DB 3677 GGCCTGTGTGCGCCACGACGCTCTTACCAACGAGTCAACGGGAGGCGCCGGCCATCGGC 3618
|
|
|
OY 541 acccttggcgctgcgcgtctgagagcaaggagcaagcgctacccctggagagctacatc 600
|
|
|
DB 3617 ACCCTTGGCGCTGCGGCTGAGAGGCCAGGAAAGCCAGCTCACCTGGAGCGATACATC 3558
|
|
|
OY 601 taccctggagagacctcgccctagacatlggcaacggcgctgtcttcccttcaactgggtc 660
|
|
|
DB 3557 TACCCCGAGAGAACCTCGCCCTAGACATGCGCAACGGCGTGTCTCTCAACTGAGGTC 3498
|
|
|
OY 661 aaggctaccagagcaacggaatccgcccccacccctgagcgctgtgtctctctgtgctcag 720
|
|
|
DB 3497 AAGGCTACACAGAGACCGGAATCCGCCACCTGAGCTGTGCTCTTGGGCTCAG 3438
|
|
|
OY 721 gggaaaaaggtgtagtcccaacacaaagacgttggcgcttaccctgggcccctatcctgtgc 780
|
|
|
DB 3437 GGGAAAAGGAGTGAAGCCCAACACCAAGACGTTGCGCTTGAACCTGGGCTCATCTGCTC 3378
|
|
|
OY 781 ctccccaagttgagcgcttccaactccgcgccttacccttaccctcattgtctacgtacat 840
|
|
|
DB 3377 CTCGCCGAGGTGAGCGCTTCCAAACCTCCGCGCCCTTATACACCTCATTTGCTAGCTACAT 3318
|
|
|
OY 841 ggcgatacctagatgacgcgttcgttcaagacgcttctatgcaagctgtgtgtggtcgtg 900
|
|
|
DB 3317 GCCGATCTCTAGATGAGCCCTGCTTCAAGAGCTTCTATGCAAGCTGTGCTGGGCTGTG 3258
|
|
|
OY 901 gccaggggtgaacccccgcgcgaatatctatttgcgttccatgagcggttatccagat 960
|
|
|
DB 3257 GCCAGGGGTGAACCTCCCGGCAATATCTATTTCCTCTAATGCGGGTTATCCGAGAT 3198
|
|
|
OY 961 taccagatgtgcaatctgacagacgcggagcgctacactgaaagacctcaagagagcc 1020
|
|
|
DB 3197 TACACGAGTGGCCATCTGACACGACGAGGAGCTACTAGTGAAGACCTCAAGAGAGGCC 3138
|
|
|
OY 1021 tctcgtga 1026
|
|
|
DB 3137 TCTCTGA 3132

```

VERSION

AL023797.1 GI:3191976

KEYWORDS

30S ribosomal protein S16; 30S ribosomal protein S2; 50S ribosomal protein L19; cell division protein FtsH homolog; elongation factor ts; ftsH; ftsH; glnB; glnD; nitrogen regulatory protein PII; prophase; protein pili uridylyltransferase; pyrH; RNA polymerase sigma factor whiG; rplS; rpsB; rpsP; signal peptidase I; signal recognition particle protein; sip1; sip2; sip3; sip4; trmD; tRNA (guanine-N1)-methyltransferase; tsf; uridylylate kinase; whiG.

SOURCE

Streptomyces coelicolor A3(2).

ORGANISM

Streptomyces coelicolor A3(2). Bacteria; Firmicutes; Actinobacteria; Actinobacteriidae; Streptomyces.

REFERENCE

1. (bases 1 to 38962)

AUTHORS

Murphy, L. and Harris, D.

TITLE

Unpublished

REFERENCE

2. (bases 1 to 38962)

AUTHORS

Parthill, J., Barrell, B.G. and Rajandream, M.A.

TITLE

Direct Submission

REFERENCE

Submitted (03-JUN-1998) Streptomyces coelicolor sequencing project, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: barrellesanger.ac.uk Cosmids supplied by Prof. David A. Hopwood. [3] John Innes Centre, Norwich Research Park, Colney, Norwich, Norfolk NR4 7UH, UK

REFERENCE

3. (bases 1 to 38962)

AUTHORS

Redenbach, M., Kleiser, H.M., Denaplatte, D., Eichner, A., Cullum, J., Kinsahl, H. and Hopwood, D.A.

TITLE

A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome

JOURNAL

Mol. Microbiol. 21 (1), 77-96 (1996)

MEDLINE

97000351

COMMENT

Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC. Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web. (URL: <http://www.sanger.ac.uk/Projects/S.coelicolor/>) CDS are numbered using the following system eg SC7B7.01c. SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand).

The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons is given for each CDS. Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon. Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot program of Bibb et al., Gene 30:157-66(1984) as implemented at <http://www.nih.gov/jp/jun/cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gta, ttg or (atc)) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid 2E1 lies between 7A1 and 6A9 on the AseI-B genomic restriction fragment.

FEATURES

1. 38962

SOURCE

/organism="Streptomyces coelicolor A3(2)"

gene

/strain="A3(2)"

CDS

/db_xref="taxon:100226"

FEATURES

/clone="cosmid 2E1"

FEATURES

1. 166

FEATURES

/gene="glnB"

FEATURES

<1. 166

FEATURES

/note="SC2E1.01, glnB, nitrogen regulatory protein pII,

partial CDS. len: >54 aa; highly similar to many e.g. GINB_AZORR (112 aa), fasta scores; opt: 198 z-score: 381.9. E(): 4.9e-14, 50.0% identity in 54 aa overlap"
/codon_start=2
/transl_table=11
/product="nitrogen regulatory protein pII"
/protein_id="CAAI9376.1"
/db_xref="GI:3191977"
/db_xref="SPTREMBL:O69872"
/translation="IRIEVLVEDDDAEOLIDVYKAAFTGKIGGKYAWVVDTRAVRRTGREGPDAL"
/gene="gind"
231. 2738
/gene="gind"
231. 2738
/gene="gind"
/note="SC2EL.02, probable gind, protein pII uridylyltransferase, len: 835 aa; similar to many e.g. GIND_ECOLI (890 aa), fasta scores; opt: 403 z-score: 540.4. E(): 7.3e-23, 26.2% identity in 851 aa overlap. Contains 7 degenerate repeats of the sequence PSS"
/codon_start=1
/transl_table=11
/product="putative protein pII uridylyltransferase"
/protein_id="CAAI9377.1"
/db_xref="GI:3191978"
/db_xref="SPTREMBL:O69873"
/translation="MTDEEDSGPGGVAAARLLITBEGASGPPARVALITPTMGYALGFAATEHNGTSLVAVGAGRGEGTSPRDLIDLDGSDGPPARVALIDRLTPYWDIGLIDHSVTPPOAKRTACODLVKHLGLIDRLABDGLTSLRTAVIADRNQDA PKRLPELDLCAREROGELQLELPDLKARAGELRATVLAARVAASWLDAREGLAEARRRLDRDCHALGTATDRLAQEOQVAAGLGLIDADALLRQYVEARVETSYADVVAERGVRLRSRSVPRLRAMNONGKGVAEKSPVLAEGVEDQEGVALPAPAREPDAIPRLRAAAAQAQGLPSLRHAAVRLAATAPLPPTPEAREPOLVTLLSGRPVQVAALEAGELVTRLLPDMERYCGRPOKNAVHMLVDHSLITVRAAGFTFRRAEDLLIALALLHDGIDKGMWDHSAGETIARDVAAKRGCDPAVATVATVIRHILLV EATSRGLDDPAPAVAAVQATCTETELLALITEADALATGPAAMSMKOSLIADILLV KRVSVLAGEEPOPEAESNAPTAEDERLVAEVRFGGVPPLAIAQTEPPADAPAPSSPSSPSPSSADGEPPELVELLIAVPPQADVLPVAVGVILAMHRLTVAATLRSVLPDVEESVLLDMRVAAQGSLEQAARLADVLTRDGLTDLIAARAEADAH PRRKVEEPPVAVAPAAASRLATVIEVRADAGGLFLGRALAEAGVRSRSHVST LGANVADFVYVTRGEGTLPDDEAAVARGESLRT"
2091. 2153
repeat_region
/note="7 degenerate 9bp repeats"
2719. 2723
/gene="gind"
/note="possible RBS upstream of ffh"
2735. 4387
/gene="ffh"
2735. 4387
/gene="ffh"
/note="SC2EL.03, ffh, signal recognition particle protein, len: 550 aa; highly similar to many e.g. SR54_ECOLI (453 aa), fasta scores; opt: 1465 z-score: 1282.8 E(): 0, 50.0% identity in 458 aa overlap. Contains PS00017 ATP/GTP-binding site motif A (P-loop), PS00300 SRP54-type proteins GTP-binding domain signature and Pfam match to entry SRP54 PF00448, SRP54-type proteins, score 66.57. Similarities suggest possible start site at aa 35"
/codon_start=1
/transl_table=11
/product="signal recognition particle protein"
/protein_id="CAAI9378.1"
/db_xref="GI:3191979"
/db_xref="SPTREMBL:O69874"
/translation="WTYRSVQRACITLPGRIPIWRTHYNADARTYAASFDTLSRLS ATFFSLRKSGKLTLEADIDATREIRALLLEADVALPVYAFIRIKVRSRSGAEVSAL NPAQOVLKIVNEELVIGGETRRLRAGAPQPTVIMLAGOAGKATLAKLSIKELIKE QHGSVLACQOELQRPNAVNOVLSVAERAGAAVYAPPEGNGOQGVKVAKSIIEFAAK VHDIVYIVTARGLGIDELMOQADRIDVAPSDLLLVVAMIGODAVNTAFIRFGV GFDQVLSKIDLDARGAALSIASVTKQPIMPASNGEKLDDPDAPFHDPRASRIIDMG DLTLLIRAEKTSQEEAEKAKSLAKKGGDFLLDDFLQMEQVVRKWSISKLIDML PGMOCKQIINLDERDVTAAITKSTGTGEKRGPTIINGSRRARIAKSGVSEAV KNLVERFEAKNMKSRNAOGGMPGMGGPPGRQKQKQKAKGQKRSQPNPKR

	misc_feature	KQOEPEAARAAAAGSGALGIPQGOGDFELPDEFKFKKG"
		/gene="ftsh"
		/note="PS00017 ATP/GTP-binding site motif A (P-loop)"
	misc_feature	3565..3697
		/gene="ftsh"
		/note="Pfam match to entry SRP54 PF00448, SRP54-type proteins, score 66.57"
	misc_feature	3647..3688
		/gene="ftsh"
		/signature="PS00300 SRP54-type proteins GTP-binding domain signature"
		4405..4453
	stem_loop	/note="hairpin loop with 22/23 bp stem"
		4504..6450
	gene	/gene="ftsh"
	CDS	4504..6450
		/gene="ftsh"
		/note="SC2E1.04, ftsh, cell division protein FtsH homolog (zinc metalloprotease, integral membrane protein), len: 648 aa; similar to many e.g. PFSH_LACIA (695 aa), fasta scores; opt: 1489 z-score: 1195.5 E(-) 0., 41.7% Identity in 640 aa overlap, PS00674 AAA-protein family signature and Pfam matches to entry AAA PF00004, ATases associated with various cellular activities (AAA) and to entry zn-proteases PF00099, zinc-binding metalloproteinase domain, score 22.07, score 357.96. Proline-rich N-terminus is not present in other ftsH homologs"
		/codon_start=1
		/transl_table=1
		/product="cell division protein FtsH homolog"
		/protein_id="CAI19379.1"
		/db_xref="GI:3191980"
		/db_xref="SPRMBL:O69875"
		/translation="MTNPSPPRKAREPWRTREGTRDPDRPKPGCGRRMRGGMNLITAA ALIYLIANLVSEFNEDPTISYTEKSVDGNVSKITYAKDAILOGLTKARDNRNP EGDGYTTEFTTERLPFADDOIMADLTNRKVTVAPVVQHSHFLANIILAAPMLIDIV VLMTFIARRMGALGGGAGGMIGRAKPRLVELBGRPRTFADVAGIDEVEGLSDVO VDLFKNPDAYRMGAKKMPRGVLLGPGETGTTLARAAGAVPFPSASAEFEEMII VGVSRSRELEVIATNRADIITFDIEDDTIGRARGGSSGGGHDERBOTNOILITEM DGFGSEGIVEIANTNRADITDALITRGRDPRVSVSPPGCRREALIEHTHEIPL APPDIIDAIVARTTTGMTCAELIANLANEAALLAYVRKDERTYOANLSLEKVOIGABR PLVMBERRRRTAFVHESGHALVLMLOPEADPVKRITTYPRRALGVTLSTPDARKVAW TEELYLRGRIIALLGMAEHVYMKITTGSSDSDEOVTNILARGVAMWSERYSRLS ALPBGQAQVGLAAPTDLIDEMERVVDSCEYAVERKLNDHRGDLDALAEEILLAS ETLDEADRYARIQTIRLIKDDPEA" 5215..5781
	misc_feature	/gene="ftsh"
		/note="Pfam match to entry AAA PF00004, ATases associated with various cellular activities (AAA), score 357.96"
Query Match		5.8%; Score 59.8; DB 1; Length 38962;
Best Local Similarity		46.3%; Pred. No. 0.22; Indels 0; Gaps 0;
Matches 196;	Conservative	0; Mismatches 227;
OY	374 ggaagtgaggacggtcgccttcattcgatcccccctggagagtgttggccctcaacctgyggg	433
Dd	1423 ggagagccttgtagagcccgagggcctgtgcatccgccctctgccccagactggagcgggtcc	1482
OY	434 tcaccggcgagacscgttcacggccttgaaagaagttccttgagaaaaagggcctgtgygcca	493
Dd	1483 gCTGCCCCCGCAGCGCANAGCCGTGCACCTGTGGACCCTGTGACCGGACCGACTGTGACGAGA	1542
OY	494 ccgaagctccttacacaacacgltcaacggggagcgcggccgatcgacaccttuggycgc	553
Dd	1543 cCGCGTCCTCGCGCCCCCGGGGTTCACCCC CGGGTGCACCGGCCCGACCTGCTCCGTGACG	1602
OY	554 tccggctgtaggcgaaggaaagccagagctcaccttggacgtaactaatcacccttgagga	613
Dd	1603 CCGCGCTCTGTCACACATCGCAAAGCGTGGCCCGGCGCACCTCCCTGTGGCCGGCAGA	1662
OY	614 acctgcctaacatctgagccaaagcctgtgatactctttaacttgtatcaaagcctaacaga	673

Db 1663 CCATGCGCCGAGACCTCGCGCCGCGCATCGCTTGACGCGCGGACACAGCGCGTCTGG 1722
 Oy 674 accacgaatccgccaccctgaacgtgctgtccttgagctcgaaggaaaggtgta 733
 Db 1723 CCACCTCGTACGACACACTCTCTGCTGTGAGACCGCCACCGCGCGGACCTGGAGC 1782
 Oy 734 tggccaaacccagacgctgacgttgacctgacctcctcctcccccagagtg 793
 Db 1783 ATCCGCGCCGCTGCGCGCGGCGCAGCGGCGGCGACGACACACCTGGAGCTGC 1842
 Oy 794 agc 796
 Db 1843 TGC 1845

RESULT 6
 SCC42/c 13347 bp DNA linear BCT 27-MAY-2000
 LOCUS Streptomyces coelicolor cosmid C42.
 DEFINITION
 AL356992.1 GI:8247035
 ABC-transporter ATP-binding protein; ABC-transporter integral
 membrane protein; phosphate binding protein; regulatory protein;
 secreted beta-galactosidase; secreted protein.
 Streptomyces coelicolor A3(2).
 Streptomyces coelicolor A3(2).
 Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 Actinomycetales; Streptomycetaceae; Streptomyces.
 1 (bases 1 to 13347)
 Kienast, H., Kleser, H.M., Denapalte, D., Elchner, A., Cullum, J.,
 Renshaw, M., and Hopwood, D.A.
 A set of ordered cosmids and a detailed genetic and physical map
 for the 8 Mb Streptomyces coelicolor A3(2) chromosome
 Mol. Microbiol. 21 (1), 77-96 (1996)
 2 (bases 1 to 13347)
 Brown, S.P. and Harris, D.
 Unpublished
 3 (bases 1 to 13347)
 Cerdano, A.M., Parkhill, J., Barrell, B.G. and Rajandream, M.A.
 Direct SubMISSION
 Submitted (26-MAY-2000) Streptomyces coelicolor sequencing project,
 Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
 CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
 David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
 Colney, Norwich, Norfolk NR4 7UH, UK

REFERENCE
 AUTHORS
 JOURNAL
 TITLE
 REFERENCE
 AUTHORS
 JOURNAL

COMMENT

Notes:
 Streptomyces coelicolor sequencing at The Sanger Centre is funded
 by the BBSRC and Beowulf Genomics
 Details of S. coelicolor sequencing at the Sanger Centre are
 available on the World Wide Web.
 (URL: <http://www.sanger.ac.uk/Projects/S.coelicolor/>)
 CDS are numbered using the following system eg SC7B7.01c. SC (S.
 coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary
 strand).
 The more significant matches with motifs in the PROSITE database
 are also included but some of these may be fortuitous.
 The length in codons is given for each CDS.
 Usually the highest scoring match found by fasta -o is given for
 CDS which show significant similarity to other CDS in the database.
 The position of possible ribosome binding site sequences are given
 where these have been used to deduce the initiation codon.
 Gene prediction is based on positional base preference in codons
 using a specially developed Hidden Markov Model (Krogh et al.,
 Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot
 program of Bibb et al., Gene 30:157-66(1984) as implemented at
<http://www.nih.go.jp/jun/cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the
 correct initiation codon. Where possible we choose an upstream
 codon (atg, gtg, ttg or (atc)) which is preceded by an upstream
 ribosome binding site sequence (optimally 5-13bp before the
 initiation codon). If this cannot be identified we choose the most

upstream initiation codon.
 IMPORTANT: This sequence MAY NOT be the entire insert of the
 sequenced clone. It may be shorter because we only sequence
 overlapping sections once, or longer, because we arrange for a
 small overlap between neighbouring submissions.
 Cosmid C42 overlaps with cosmid C24 on the Asel-C.
 location/Qualifiers

FEATURES

source

misc-feature

CDS

complement(1..486)
 /gene="SCC42.01c", possible ABC-transporter integral
 membrane protein (fragment), len: >161 aa; similar to
 TR:053545 (EMBL:AL022022) Mycobacterium tuberculosis
 hypothetical 29.6 kD protein MY023.07c, 280 aa; fasta
 scores: opt: 565 z-score: 630.0 E(): 1.3e-27; 54.58
 identity in 154 aa overlap. Contains possible hydrophobic
 membrane spanning regions"
 /codon_start=1
 /transl_table=11
 /product="putative ABC-transporter integral membrane
 protein (fragment)"
 /protein_id="CAB92894.1"
 /db_xref="GI:8247036"
 /translation="MASPLVWDRSGDQLFLYRALLMVPRTLRRLKEVORLAEVA
 FSGGLGVIGTIGVMIAMLTFTVGIGVLAADIGTSAFGEVSAVFETRIAP
 LVAGIALATVAGFTAQDAGMRINEEDALEGMGRMPYLVTTRIIAGVAILPL
 A1"
 complement(1..486)
 /gene="SCC42.01c"
 complement(491..494)
 complement(493..1257)
 /gene="SCC42.02c"
 complement(493..1257)
 /gene="SCC42.02c"
 /note="SCC42.02c, possible ABC-transporter integral
 membrane protein, len: 254 aa; similar to TR:053546
 (EMBL:AL022022) Mycobacterium tuberculosis hypothetical
 26.6 kD protein MY023.08c, 254 aa; fasta scores: opt: 881
 z-score: 1004.0 E(): 0; 50.88 identity in 242 aa overlap.
 Contains possible hydrophobic membrane spanning regions"
 /codon_start=1
 /transl_table=11
 /product="putative ABC-transporter integral membrane
 protein"
 /protein_id="CAB92895.1"
 /db_xref="GI:8247037"
 /translation="MTTGLRGTGRFLAAEVRVPRPREFVROFVAVST
 ILPALVSTIPFGAVIALQVSLTEDGASQFTFGASVLAAYQAAPLIYALIAAGG
 SAICADLGSRRIRLEIDAEVMGVSVPQVVPVRYLAAGVALLAGLVSVGIIGIG
 FGVNIWOGCTPGAYLSFSAALQDPLVYSEIKALVFGIAGIIVAAVRLNDRGPKG
 VGDVAVOSVITFLILFVNMWMTAVYLIQIVPKG"
 complement(1254..2285)
 /gene="SCC42.03c"
 complement(1254..2285)
 /gene="SCC42.03c"
 /note="SCC42.03c, possible ABC-transporter ATP-binding
 protein, len: 343 aa; similar to SW:YRBF_HAEIN
 (EMBL:U32788) Haemophilus influenzae hypothetical
 ABC-transporter ATP-binding protein H11087, 264 aa; fasta
 scores: opt: 680 z-score: 686.6 E(): 8.9e-31; 41.4%
 identity in 244 aa overlap. Contains Pfam match to entry
 PF00005 ABC_tran, ABC transporter and matches to Prosite
 entries PS00017 ATP/GTP-binding site motif A (P-loop) and
 PS00211 ABC transporters family signature"
 /codon_start=1
 /transl_table=11

/product="putative ABC-transporter ATP-binding protein"
/protein_id="CAB92896.1"
/db_xref="GI:8247038"
/translation="MASAGDMGIEVVEGLTKSFQKQIMRDVSLTLPAEVSVMGL
PSGKIVFELKSLGLPEKGRVLIINCVNNSPERDIYERKILGIMFGDGLFSG
MSLEPNIAFLREHREKRESEIRIYMRIVNIVYGLAEKLPKISGMRKGLAR
ALVLDPOIILCDEPDSGLDPVTAIISOLIDLNAQIDATMLIYTHNIDIAIYDNN
GMLFLROLVTEGPREVSLTSDPVAOPIGGRRCRPIGMSSEKAATIAADAPAP
PAAPRVITPOLEPSGLPPRAVARRRRKRVGMIDTLLPAAKSAIRDTYAROSEATL
PMPAAGSCA"
RBS
complement(1264..1267)
/gene="SCC42.03c"
complement(1620..2177)
/gene="SCC42.03c"
/note="Pfam match to entry PF00005 ABC_tran, ABC
transporter, score 152.90, E-value 5.4e-42"
misc-feature
complement(1800..1844)
/gene="SCC42.03c"
/note="PS00211 ABC transporters family signature"
complement(2133..2156)
/gene="SCC42.03c"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"
2904..2909
2918..3350
/gene="SCC42.04"
2918..3350
/gene="SCC42.04"
/note="SCC42.04"
Contains possible N-terminal region signal peptide
sequence
/codon_start=1
/transl_table=11
/product="putative secreted protein"
/protein_id="CAB92897.1"
/db_xref="GI:8247039"
/translation="MKKNMRSLALVYCAASALGLAAPPSSAVPSTVWVNPSPAAK
AVANGNIVLRITPTCTTSTAAGTMASAGNPGDVAIDITMNGASGSPCTSYLGNV
ITTSVTWVDVGVNDSITGVTGVNVAIVAGACKFTYNGKASGITNSTGVLA
INSTAGELAVSNPNCALIVTATKPTFKGVAVTVASTGAIPITVSNP"
RBS
3679..3682
3687..5288
/gene="SCC42.05"
3687..5288
/gene="SCC42.05"
/note="SCC42.05"
Contains possible N-terminal region signal peptide
sequence. High content in alanine amino acid residues"
/codon_start=1
/transl_table=11
/product="putative secreted protein"
/protein_id="CAB92898.1"
/db_xref="GI:8247040"
/translation="MKATPAAPRPSRARATTSIAFVILALVPTTASAGTOEVE
AELPYCTLPSCGLPATVRVSAEPERAGADEFTPDVYTVLEIAPAEVATADDA
AELRAATSLAVGAONATATVATWGRASAEVALPGSRPLTVRGVPSVAGSDGL
TFASGALADIALAGADPATADPSGLTVDCTLLDEAGGLATVPEVGTGQAPSSP
SSSGPAGSSGAPDDGKQDGGDRRPEROSERPKYLENPGAAARDVDVPCRYDEO
HPPLVDLSNAVIGDVKWKKMGAAVLPSCVILEGLPVPGPDEYLLFDLSYAN
FHYRERKQTPFEATLSPDFEAVKATVLEQGTMRKIDSRMRLSDFTIDTVAN
APLVILDLVNGTPLDVGSECRTERETSLSEPDNDPVDHLYVYGREOLIGIPA
TGYLLSGALGSEATIPATFGCGSDGDDRLTLTASVSPGVINQVQCQICAIAP
VFPSPNEGCTCDIQPELPAER"
gene
5323..6264
/gene="SCC42.06"
5323..6264
/gene="SCC42.06"
/note="SCC42.06"
Contains possible N-terminal region signal peptide
sequence"
/codon_start=1
/transl_table=11
/product="putative secreted protein"

Query Match		5.5%	Score 56.8	DB 1	Length 13347
Best Local Similarity		45.6%	Pred. No. 0.89		
Matches 199		Conservative	0	Mismatches 237	Indels 0
					Gaps 0
Qy	339	ccccacatccccctccgcaaccggaagctctctgaggaagtgaggagcgtcttcat	398		
Db	724	CTCCCCCTCCGCCACGCTGCCGACCTTACGTCAAGCAAGCGCTGCTTGG	665		
Qy	399	gtcccccttgagatgttgcctcaaccttgagggtacaccgacagctccagcctg	458		
Db	664	GTTCAATCCGGGCAATCGTCGCCGCTACCGGGAATCAACCGCGCGCCCAACGG	605		
Qy	459	gaagaagctcttgagaaagggcctgtgtccacagagctcttcaaccaacctga	518		
Db	604	CGTCGCCACCCGCAACCAAGTCGTCATCACTTCCTGCTTCCTTCGTCAA	545		
Qy	519	cggagagcggcgacatcggacaccttggccgtccgtcgttgagccaggaagcag	578		
Db	544	CATGCTATGACGGCGGCTACCTCAATGTCGCCGGAAGGAGGTAGTCCGAT	485		
Qy	579	gtcaacctgagactacatcacctcctgaggaacctcgccctagacatgccaacg	638		
Db	484	GGCTCCCGCTCGCTGCTGCTGACCGCTCGGCGAACCACTGCTTCTACGTCGGGC	425		
Qy	639	cgctctccttaactggtgctcaagcgcttaccagagacaggaatcgccccctgga	698		
Db	424	CCTGCTGGGTCCCGCGGACCTGCGCGCTACCTCAAGAGGTGACGCGCTCTCGC	365		
Qy	699	cgctgctcctctggtgctcaggaagagtgatgcccacacaaagacgtgycgt	758		
Db	364	CGAGGTGGCTTTCGGCTCGCGCGCTCGGCTCATCGCGGCAACATCGCGCTGATGAT	305		
Qy	759	tgacctgggctcatc 774			
Db	304	CGCATGACGCTCTTC 289			
RESULT 7					
LOCUS	SC2D46	35576 bp	DNA	linear	BC1 18-AUG-2000
DEFINITION	Streptomyces coelicolor cosmid 2D46.				
ACCESSION	AL391406				
VERSION	AL391406.1 GI:9857143				
KEYWORDS	AbaA-like; Arac-family; ATP/GTP-binding membrane; bldA; DNA-binding; GntR-family transcriptional regulator; hydrolase; integral; integrase; MarR-family; NRP/P60-family secreted; pgu; phosphoglycerate mutase; TelR-family; transferase; transport. Streptomyces coelicolor A3(2). Streptomyces coelicolor A3(2). Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.				
SOURCE	ORGANISM				
REFERENCE	Redenbach,M., Kleiser,H.M., Denapalte,D., Eichner,A., Cullum,J., Kinasli,H. and Hopwood,D.A. A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome 97000351				
AUTHORS	Brown,S.P. and Harris,D.				
REFERENCE	Bentley,S.D., Parkhill,J., Barrell,B.G. and Randsdram,M.A. Submitted (03-AUG-2000) Streptomyces coelicolor sequencing project, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof. David A. Hopwood, [3] John Innes Centre, Norwich Research Park, Colney, Norwich, Norfolk NR4 7UH, UK				
AUTHORS	Notes:				
TITLE	Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC and Beowulf Genomics				
JOURNAL	Details of S. coelicolor sequencing at the Sanger Centre are				

available on the World Wide Web.
(URL: http://www.sanger.ac.uk/Projects/S_coelicolor/)
CDS are numbered using the following system eg SC7B7.01c, SC (S, coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand).

The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous.
The length in codons is given for each CDS.

Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon.

Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the Frameplot program of Bibb et al., Gene 30:157-66(1984) as implemented at <http://www.nih.go.jp/jun/cgi-bin/frameplot.pl>.

CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or (att)) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.

Cosmid 2D46 lies between and overlaps cosmids D66 and D8 on the AseI-D genomic restriction fragment.

FEATURES

source

1. 35576

/location/Qualifiers

1. 35576

/organism="Streptomyces coelicolor A3(2)"

/strain="A3(2)"

/db_xref="taxon:100226"

/clone="cosmid 2D46"

complement(1..135)

/gene="2SCD46.01c"

1..1200

/note="Repeated degenerately at 1210..2414. This DNA repeat manifests as a duplication of a pairs of CDSs such that 2SCD46.01c is similar to 2SCD46.03c and 2SCD46.02 is similar to 2SCD46.04."

complement(1..135)

/gene="2SCD46.01c"

1..1200

/note="Repeated degenerately at 1210..2414. This DNA repeat manifests as a duplication of a pairs of CDSs such that 2SCD46.01c is similar to 2SCD46.03c and 2SCD46.02 is similar to 2SCD46.04."

complement(1..135)

/gene="2SCD46.01c"

1..1200

/note="Repeated degenerately at 1210..2414. This DNA repeat manifests as a duplication of a pairs of CDSs such that 2SCD46.01c is similar to 2SCD46.03c and 2SCD46.02 is similar to 2SCD46.04."

complement(1..135)

/gene="2SCD46.01c"

1..1200

/note="Repeated degenerately at 1210..2414. This DNA repeat manifests as a duplication of a pairs of CDSs such that 2SCD46.01c is similar to 2SCD46.03c and 2SCD46.02 is similar to 2SCD46.04."

complement(1..135)

/gene="2SCD46.01c"

1..1200

/note="Repeated degenerately at 1210..2414. This DNA repeat manifests as a duplication of a pairs of CDSs such that 2SCD46.01c is similar to 2SCD46.03c and 2SCD46.02 is similar to 2SCD46.04."

complement(1..135)

/gene="2SCD46.01c"

1..1200

/note="Repeated degenerately at 1210..2414. This DNA repeat manifests as a duplication of a pairs of CDSs such that 2SCD46.01c is similar to 2SCD46.03c and 2SCD46.02 is similar to 2SCD46.04."

complement(1..135)

/gene="2SCD46.01c"

1..1200

/note="Repeated degenerately at 1210..2414. This DNA repeat manifests as a duplication of a pairs of CDSs such that 2SCD46.01c is similar to 2SCD46.03c and 2SCD46.02 is similar to 2SCD46.04."

complement(1..135)

/gene="2SCD46.01c"

1..1200

/note="Repeated degenerately at 1210..2414. This DNA repeat manifests as a duplication of a pairs of CDSs such that 2SCD46.01c is similar to 2SCD46.03c and 2SCD46.02 is similar to 2SCD46.04."

complement(1..135)

/gene="2SCD46.01c"

1..1200

/note="Repeated degenerately at 1210..2414. This DNA repeat manifests as a duplication of a pairs of CDSs such that 2SCD46.01c is similar to 2SCD46.03c and 2SCD46.02 is similar to 2SCD46.04."

complement(1..135)

/gene="2SCD46.01c"

1..1200

/note="Repeated degenerately at 1210..2414. This DNA repeat manifests as a duplication of a pairs of CDSs such that 2SCD46.01c is similar to 2SCD46.03c and 2SCD46.02 is similar to 2SCD46.04."

complement(1..135)

/gene="2SCD46.01c"

1..1200

/note="Repeated degenerately at 1210..2414. This DNA repeat manifests as a duplication of a pairs of CDSs such that 2SCD46.01c is similar to 2SCD46.03c and 2SCD46.02 is similar to 2SCD46.04."

complement(1..135)

/gene="2SCD46.01c"

1..1200

/note="Repeated degenerately at 1210..2414. This DNA repeat manifests as a duplication of a pairs of CDSs such that 2SCD46.01c is similar to 2SCD46.03c and 2SCD46.02 is similar to 2SCD46.04."

complement(1..135)

/gene="2SCD46.01c"

1..1200

/note="Repeated degenerately at 1210..2414. This DNA repeat manifests as a duplication of a pairs of CDSs such that 2SCD46.01c is similar to 2SCD46.03c and 2SCD46.02 is similar to 2SCD46.04."

complement(1..135)

/gene="2SCD46.01c"

1..1200

/note="Repeated degenerately at 1210..2414. This DNA repeat manifests as a duplication of a pairs of CDSs such that 2SCD46.01c is similar to 2SCD46.03c and 2SCD46.02 is similar to 2SCD46.04."

complement(1..135)

/gene="2SCD46.01c"

1..1200

/note="Repeated degenerately at 1210..2414. This DNA repeat manifests as a duplication of a pairs of CDSs such that 2SCD46.01c is similar to 2SCD46.03c and 2SCD46.02 is similar to 2SCD46.04."

complement(1..135)

/gene="2SCD46.01c"

1..1200

/note="Repeated degenerately at 1210..2414. This DNA repeat manifests as a duplication of a pairs of CDSs such that 2SCD46.01c is similar to 2SCD46.03c and 2SCD46.02 is similar to 2SCD46.04."

complement(1..135)

/gene="2SCD46.01c"

1..1200

/note="Repeated degenerately at 1210..2414. This DNA repeat manifests as a duplication of a pairs of CDSs such that 2SCD46.01c is similar to 2SCD46.03c and 2SCD46.02 is similar to 2SCD46.04."

complement(1..135)

/gene="2SCD46.01c"

1..1200

/note="Repeated degenerately at 1210..2414. This DNA repeat manifests as a duplication of a pairs of CDSs such that 2SCD46.01c is similar to 2SCD46.03c and 2SCD46.02 is similar to 2SCD46.04."

complement(1..135)

/gene="2SCD46.01c"

1..1200

/note="Repeated degenerately at 1210..2414. This DNA repeat manifests as a duplication of a pairs of CDSs such that 2SCD46.01c is similar to 2SCD46.03c and 2SCD46.02 is similar to 2SCD46.04."

complement(1..135)

/gene="2SCD46.01c"

1..1200

/note="Repeated degenerately at 1210..2414. This DNA repeat manifests as a duplication of a pairs of CDSs such that 2SCD46.01c is similar to 2SCD46.03c and 2SCD46.02 is similar to 2SCD46.04."

complement(1..135)

/gene="2SCD46.01c"

1..1200

/note="Repeated degenerately at 1210..2414. This DNA repeat manifests as a duplication of a pairs of CDSs such that 2SCD46.01c is similar to 2SCD46.03c and 2SCD46.02 is similar to 2SCD46.04."

complement(1..135)

/gene="2SCD46.01c"

1..1200

/note="Repeated degenerately at 1210..2414. This DNA repeat manifests as a duplication of a pairs of CDSs such that 2SCD46.01c is similar to 2SCD46.03c and 2SCD46.02 is similar to 2SCD46.04."

complement(1..135)

/gene="2SCD46.01c"

1..1200

/note="Repeated degenerately at 1210..2414. This DNA repeat manifests as a duplication of a pairs of CDSs such that 2SCD46.01c is similar to 2SCD46.03c and 2SCD46.02 is similar to 2SCD46.04."

complement(1..135)

/gene="2SCD46.01c"

1..1200

/note="Repeated degenerately at 1210..2414. This DNA repeat manifests as a duplication of a pairs of CDSs such that 2SCD46.01c is similar to 2SCD46.03c and 2SCD46.02 is similar to 2SCD46.04."

complement(1..135)

/gene="2SCD46.01c"

1..1200

/note="Repeated degenerately at 1210..2414. This DNA repeat manifests as a duplication of a pairs of CDSs such that 2SCD46.01c is similar to 2SCD46.03c and 2SCD46.02 is similar to 2SCD46.04."

complement(1..135)

/gene="2SCD46.01c"

1..1200

/note="Repeated degenerately at 1210..2414. This DNA repeat manifests as a duplication of a pairs of CDSs such that 2SCD46.01c is similar to 2SCD46.03c and 2SCD46.02 is similar to 2SCD46.04."

complement(1..135)

/gene="2SCD46.01c"

1..1200

/note="Repeated degenerately at 1210..2414. This DNA repeat manifests as a duplication of a pairs of CDSs such that 2SCD46.01c is similar to 2SCD46.03c and 2SCD46.02 is similar to 2SCD46.04."

complement(1..135)

/gene="2SCD46.01c"

1..1200

/note="Repeated degenerately at 1210..2414. This DNA repeat manifests as a duplication of a pairs of CDSs such that 2SCD46.01c is similar to 2SCD46.03c and 2SCD46.02 is similar to 2SCD46.04."

complement(1..135)

/gene="2SCD46.01c"

1..1200

/note="Repeated degenerately at 1210..2414. This DNA repeat manifests as a duplication of a pairs of CDSs such that 2SCD46.01c is similar to 2SCD46.03c and 2SCD46.02 is similar to 2SCD46.04."

complement(1..135)

/gene="2SCD46.01c"

1..1200

/note="Repeated degenerately at 1210..2414. This DNA repeat manifests as a duplication of a pairs of CDSs such that 2SCD46.01c is similar to 2SCD46.03c and 2SCD46.02 is similar to 2SCD46.04."

complement(1..135)

/gene="2SCD46.01c"

1..1200

/note="Repeated degenerately at 1210..2414. This DNA repeat manifests as a duplication of a pairs of CDSs such that 2SCD46.01c is similar to 2SCD46.03c and 2SCD46.02 is similar to 2SCD46.04."

complement(1..135)

/gene="2SCD46.01c"

1..1200

/note="Repeated degenerately at 1210..2414. This DNA repeat manifests as a duplication of a pairs of CDSs such that 2SCD46.01c is similar to 2SCD46.03c and 2SCD46.02 is similar to 2SCD46.04."

complement(1..135)

/gene="2SCD46.01c"

1..1200

/note="Repeated degenerately at 1210..2414. This DNA repeat manifests as a duplication of a pairs of CDSs such that 2SCD46.01c is similar to 2SCD46.03c and 2SCD46.02 is similar to 2SCD46.04."

complement(1..135)

/gene="2SCD46.01c"

1..1200

/note="Repeated degenerately at 1210..2414. This DNA repeat manifests as a duplication of a pairs of CDSs such that 2SCD46.01c is similar to 2SCD46.03c and 2SCD46.02 is similar to 2SCD46.04."

complement(1..135)

/gene="2SCD46.01c"

1..1200

/note="Repeated degenerately at 1210..2414. This DNA repeat manifests as a duplication of a pairs of CDSs such that 2SCD46.01c is similar to 2SCD46.03c and 2SCD46.02 is similar to 2SCD46.04."

complement(1..135)

/gene="2SCD46.01c"

1..1200

/note="Repeated degenerately at 1210..2414. This DNA repeat manifests as a duplication of a pairs of CDSs such that 2SCD46.01c is similar to 2SCD46.03c and 2SCD46.02 is similar to 2SCD46.04."

complement(1..135)

/gene="2SCD46.01c"

1..1200

/note="Repeated degenerately at 1210..2414. This DNA repeat manifests as a duplication of a pairs of CDSs such that 2SCD46.01c is similar to 2SCD46.03c and 2SCD46.02 is similar to 2SCD46.04."

complement(1..135)

/gene="2SCD46.01c"

1..1200

/note="Repeated degenerately at 1210..2414. This DNA repeat manifests as a duplication of a pairs of CDSs such that 2SCD46.01c is similar to 2SCD46.03c and 2SCD46.02 is similar to 2SCD46.04."

complement(1..135)

/gene="2SCD46.01c"

1..1200

/note="Repeated degenerately at 1210..2414. This DNA repeat manifests as a duplication of a pairs of CDSs such that 2SCD46.01c is similar to 2SCD46.03c and 2SCD46.02 is similar to 2SCD46.04."

complement(1..135)

/gene="2SCD46.01c"

1..1200

/note="Repeated degenerately at 1210..2414. This DNA repeat manifests as a duplication of a pairs of CDSs such that 2SCD46.01c is similar to 2SCD46.03c and 2SCD46.02 is similar to 2SCD46.04."

complement(1..135)

/gene="2SCD46.01c"

1..1200

/note="Repeated degenerately at 1210..2414. This DNA repeat manifests as a duplication of a pairs of CDSs such that 2SCD46.01c is similar to 2SCD46.03c and 2SCD46.02 is similar to 2SCD46.04."

complement(1..135)

/gene="2SCD46.01c"

1..1200

/note="Repeated degenerately at 1210..2414. This DNA repeat manifests as a duplication of a pairs of CDSs such that 2SCD46.01c is similar to 2SCD46.03c and 2SCD46.02 is similar to 2SCD46.04."

complement(1..135)

/gene="2SCD46.01c"

1..1200

/note="Repeated degenerately at 1210..2414. This DNA repeat manifests as a duplication of a pairs of CDSs such that 2SCD46.01c is similar to 2SCD46.03c and 2SCD46.02 is similar to 2SCD46.04."

complement(1..135)

/gene="2SCD46.01c"

1..1200

/note="Repeated degenerately at 1210..2414. This DNA repeat manifests as a duplication of a pairs of CDSs such that 2SCD46.01c is similar to 2SCD46.03c and 2SCD46.02 is similar to 2SCD46.04."

complement(1..135)

/gene="2SCD46.01c"

	Bacteria; Firmicutes; Actinobacteriae; Actinobacterales; Actinomycetales; Streptomycineae; Streptomycesetaceae; Streptomyces
REFERENCE AUTHORS	1 (bases 1 to 41622) Kieserich,M., Kieseir,H.M., Denapalte,D., Eichner,A., Cullum,J., Rensbach,H. and Hopwood,D.A.
TITLE	A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome
JOURNAL	Mol. Microbiol. 21 (1), 77-96 (1996)
MEDLINE	97000351
REFERENCE	2 (bases 1 to 41622)
AUTHORS	Brown,S.P. and Harris,D. Unpublished
JOURNAL	3 (bases 1 to 41622)
AUTHORS	Thomson,N.R., Parkhill,J., Barrell,B.G. and Randsdram,M.A. Direct Submission
TITLE	Submitted (20-SEP-1999) Streptomyces coelicolor sequencing project sanger centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof David A. Hopwood, [3] John Innes Centre, Norwich Research Park, Colney, Norwich, Norfolk NR4 7QH, UK
COMMENT	Notes:

Best Local Similarity	46.6%	Pred. No. 1.9;	
Matches 174;	Conservative	0;	Mismatches 199; Indels 0; Gaps 0;

COMMENTS

Notes:

Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC and Beowulf Genomics

Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web.

(URL: <http://www.sanger.ac.uk/Projects/S.coelicolor/>)

CDS are numbered using the following system eg SC7B1.01c. SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand).

The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous.

The length in codons is given for each CDS.

Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database.

The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon.

Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the Framerplot program of Bibb et al., Gene 50:157-66(1984) as implemented at <http://www.ihh.gu.jp/jun/cgi-bin/framerplot.pl>.

CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an upstream codon (atg, gtg, ttg or attt) which is preceded by an upstream ribosome binding site sequence (optimally 5-12bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.

Cosmid D25 overlaps with cosmid D17 on the Ascl-D genomic restriction fragment.

```

FEATURES
source
    Location/Qualifiers
        1..41622
        /organism="Streptomyces coelicolor A3(2)"
        /strain="A3(2)"
        /db_xref="taxon:100226"
        /clone="cosmid D25"
        complement(1..451)
        /gene="SCD25.01c"
        complement(<1..451)
        /gene="SCD25.01c"
        /note="SCD25.01c, possible transposase, partial CDS, len: 2150 aa. Similar to many e.g. Methyllobacterium sp TR-050400 (EMBL; X96995) transposase IS157 (411 aa), fasta scores opt: 239.2 score: 312.5 E): 4.7e-10 38.2% identity in 131 aa overlap."
        /codon_start=1
        /transl_table=11
        /product="putative transposase"
        /protein_id="CAB56345.1"
        /db_xref="GI:5918467"

```


LOCUS AC105260 100772 bp DNA linear HTG 28-DEC-2001
DEFINITION *Oryza sativa* chromosome 5 clone OJ115D04, *** SEQUENCING IN
PROGRESS ***, 2 ordered pieces.
ACCESSION AC105260
VERSION AC105260.1 GI:17985870
KEYWORDS HTG: HTGS_PHASE2.
SOURCE *Oryza sativa*.
ORGANISM *Oryza sativa*.
REFERENCE Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Eriharoldae; Oryzaceae; *Oryza*.
1 (bases 1 to 100772)
Chow, T.-Y., Hsling, Y.-I.C., Chen, C.-S., Chen, H.-H., Wu, H.-P.,
Liu, S.-M., Chao, Y.-T., Chang, S.-J., Chen, T.-R., Chen, Y.-L.,
Kau, P.-I., Lee, M.-C., Leu, H.-L., Lin, S.-J., Wu, L.-F., Huang, J.-J.,
F. Shaw, J.-F.
Oryza sativa BAC OJ115D04 genomic sequence
Unpublished
2 (bases 1 to 100772)
Chow, T.-Y. and Hsling, Y.-I.C.
Direct Submission
Submitted (28-DEC-2001) Institute of Botany, Academia Sinica, 128,
Section 2, Academia Road, Nankang, Taipei 11529, Taiwan
The nucleotide sequence of this BAC clone was generated by
combining Monsanto and ASPG-Taiwan sequencing data.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 39588: contig of 39588 bp in length
* 39589 39608: gap of unknown length
* 39609 100772: contig of 61164 bp in length.
Location/Qualifiers
1. 100772
/organism="Oryza sativa"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/chromosome="5"
/clone="OJ115D04"

BASE COUNT 28103 a 22456 c 22621 g 27572 t 20 others
ORIGIN
Query Match 5.2%: Score 53.2; DB 2; Length 100772;
Best Local Similarity 50.8%; Pred. No. 2.9;
Matches 127; Conservative 0; Mismatches 123; Indels 0; Gaps 0;
OY 543 ccttggcgcgtcgctgagccaggaagccagctccctggagctatctta 602
DB 88516 CGTCTTCATGGCCGGAGCTCCGCCAGACGACGACATGACCTAGGGGTCTC 88575
OY 603 cccctggaggaaccccgccctagacatgagcagcgctgctctcctcaactgagta 662
DB 88576 CCACGGCAACGCTCTCCACCTGTCATCCGATCTCCGACCTCCGCTCATACCGTCA 88635
OY 663 ggcctacagagaccaggaatcgcgcccaaccctgagcgtgtgtctctctgggtcagg 722
DB 88636 GACCGTCCACGGCACAGATTGACGCTCCGCTCGAGCCCGCCACCGTGGCTACGT 88695
OY 723 gaaagagtgatgacccaacccaagaagcgtgagcgttgacccgggcccattctgttct 782
DB 88696 CAAAGACGATGCGCAAGAACTCCACCCAGCAGACACCAAGCTCTGCTCTCA 88755
OY 783 ccccgaggtg 792
DB 88756 GGGCGAGGTG 88765

RESULT 10
AP004746/c
LOCUS AP004746 167049 bp DNA linear HTG 14-FEB-2002
DEFINITION *Oryza sativa* chromosome 8 clone P0035F08, *** SEQUENCING IN
PROGRESS ***, in ordered pieces.
ACCESSION AP004746
VERSION AP004746.1 GI:18656392
KEYWORDS HTG: HTGS_PHASE2.
SOURCE *Oryza sativa*.
ORGANISM *Oryza sativa* (cultivar:Nipponbare) DNA, clone:P0035F08.
REFERENCE Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Eriharoldae; Oryzaceae; *Oryza*.
1 (bases 1 to 167049)
Sasaki, T., Matsumoto, T. and Yamamoto, K.
Direct Submission
Submitted (13-FEB-2002) Takuji Sasaki, National Institute of
Agrobiological Sciences, Rice Genome Research Program, Kannondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail:tsasaki@nias.affrc.go.jp, URL:http://rjp.dna.affrc.go.jp/,
Tel:81-298-38-7441, Fax:81-298-38-7468)
NOTE: It currently consists of 1 contigs. Gaps between the contigs
are represented as runs of N. The order of the pieces between them
to be correct as given, however the sizes of the gaps between them
are based on estimates that have provided by the submitter. This
sequence will be replaced by the finished sequence as soon as it is
available and the accession number will be preserved.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
Location/Qualifiers
1. 167049
/organism="Oryza sativa"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/chromosome="8"
/clone="P0035F08"

BASE COUNT 45406 a 36605 c 37338 g 47383 t 317 others
ORIGIN
Query Match 5.2%: Score 53.2; DB 2; Length 167049;
Best Local Similarity 46.3%; Pred. No. 2.7; 243; Indels 3; Gaps 1;
Matches 212; Conservative 0; Mismatches 243; Indels 3; Gaps 1;
OY 356 gcaaccggagctcgaggaaggtgggagcgtcttcatgtccctggagatgt 415
DB 166796 GCGCATGCTCTACACGGATGAAGAGAGCGCCGAGGCTTACTCGCAGAAAGTCA 166737
OY 416 tggccctcaacttggggttcacccggcagaccgtccacgcttgagaaggtcttgaga 475
DB 166736 CCGCGGCGGTGTCACCTCCCGGCTTCAACAGCAGCGGCGGCGGCGGCGGCGG 166677
OY 476 aaaaagccttggggttcacccggcagaccgtccacgcttgagaaggtcttgaga 535
DB 166676 ACGCGGGGTATCCCGCGGCTCACCGCTGACCGCATATCAACGACGCCACCGCGCG 166617
OY 536 tcggacccttggggttcacccggcagaccgttcacgcttgagaaggttcacccctgagact 595
DB 166616 CCATGCGCTTACGGCATCACAAGAGGCGCGGAGAAAGAGTCTCTCTTCAAGCTCG 166557
OY 596 acaatcaccccttggggttcacccggcagaccgttcacgcttgagaaggttcacccctcaact 655
DB 166556 GCGGCGGACGTTGACATGACATCTCGCATGACAGGCGGTCTTCAAGGTCCTTG 166497
OY 656 ggggttaagccttaccaggaacgaatcgg---cccaccccttggggttcgtctct 712
DB 166496 CCACCAACGGGACACCCACCTCGGCGGCGGAGATTGACCAACGCTCATGGACACT 166437
OY 713 gggctcaggggaaaggtgatgtcccaacacaaagaccgttgcgttgaccttgagctca 772
DB 166436 TCGTCAAGGTATCCGCGGAGAGACAGCGGCGGCGGCGGCGGCGGCTGG 166377

QY 773 tccgtgctcccccagagtgagcgcttccaactcccg 810
DB 166376 GCAGAGCTCGCGCGAGTGCAGCGCGCCAGCGCGCG
RESULT 11
SC8A2/c 7459 bp DNA linear BCT 09-OCT-2000
LOCUS Streptomyces coelicolor cosmid 8A2.
DEFINITION AL445327
ACCESSION AL445327.1 GI:10798655
VERSION Streptomyces coelicolor A3(2).
KEYWORDS secreted protein.
SOURCE Streptomyces coelicolor A3(2).
ORGANISM Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
1 (bases 1 to 7459)
Rendenbach, M., Kieser, H.M., Denapalte, D., Eichner, A., Cullum, J.,
Kinash, H. and Hopwood, D.A.
A set of ordered cosmids and a detailed genetic and physical map
for the 8 Mb Streptomyces coelicolor A3(2) chromosome
Mol. Microbiol. 21 (1), 77-96 (1996)
97000351
2 (bases 1 to 7459)
Brown, S.P. and Harris, D.
Unpublished
3 (bases 1 to 7459)
Cerdano, A.M., Parkhill, J., Barrell, B.G. and Rajandream, M.A.
Direct Submission
Submitted (09-OCT-2000) Streptomyces coelicolor sequencing project,
Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
Colney, Norwich, Norfolk NR4 7UH, UK
Notes:
Streptomyces coelicolor sequencing at The Sanger Centre is funded
by the BBSRC and Beecham Genomics
Details of S. coelicolor sequencing at the Sanger Centre are
available on the World Wide Web.
(URL: <http://www.sanger.ac.uk/Projects/S.coelicolor/>) CDS are
numbered using the following system eg SC7B7.01c. SC (S.
coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary
strand).
The more significant matches with motifs in the PROSITE database
are also included but some of these may be fortuitous. The length
in codons is given for each CDS.
Usually the highest scoring match found by fasta -o is given for
CDS which show significant similarity to other CDS in the database.
The position of possible ribosome binding site sequences are given
where these have been used to deduce the initiation codon. Gene
prediction is based on positional base preference in codons using a
specially developed Hidden Markov Model (Krogh et al., Nucleic
Acids Research, 22(22):4768-4778(1994)) and the Frameplot program
of Babb et al., Gene 30:157-66(1984) as implemented at
<http://www.nih.gov.jp/>
jun/cgi-bin/frameplot.pl. CAUTION: We may not have predicted the
correct initiation codon. Where possible we choose an initiation
codon (atg, gtg, ttg or (atc) which is preceded by an upstream
ribosome binding site sequence (optimally 5-13bp before the
initiation codon). If this cannot be identified we choose the most
upstream initiation codon.
IMPORTANT: This sequence MAY NOT be the entire insert of the
sequenced clone. It may be shorter because we only sequence
overlapping sections once, or longer, because we arrange for a
small overlap between neighbouring submissions. Cosmid 8A2 lies
between and overlaps cosmids 4A7 and C42 on the AseI-C genomic
restriction fragment.

FEATURES
Source
1..7459
/organism="Streptomyces coelicolor A3(2)"
/strain="A3(2)"
/db_xref="taxon:100226"

gene
misc_feature
CDS
gene
CDS
misc_feature
gene
misc_feature
RBS
misc_feature
/clone="cosmid 8A2"
complement(1..126)
/gene="SC8A2.01c"
1..112
/note="nominal overlap with Streptomyces coelicolor cosmid
SC4A7"
complement(1..126)
/gene="SC8A2.01c"
/note="SC8A2.01c, possible secreted protein (fragment),
len: >42 aa. Contains possible N-terminal region signal
peptide sequence"
/codon_start=1
/transl_table=11
/product="putative secreted protein (fragment)"
/protein_id="CAC12792.1"
/db_xref="GI:10798656"
/translation="MKRRRIILAGALVLAAGVCGTGMATYTGATDGLIAGYRERD"
complement(123..1424)
/gene="SC8A2.02c"
complement(123..1424)
/gene="SC8A2.02c"
/note="SC8A2.02c, possible secreted protein, len: 433 aa;
similar to TR:053972 (EMBL:AL022073) Mycobacterium
tuberculosis hypothetical 46.7 kDa protein MT051.09, 437
aa; fasta scores: opt: 529 z-score: 488.9 E(): 1e-19;
31.3% identity in 438 aa overlap. Contains Pfam match to
entry PF02470 mce. Also contains possible N-terminal
region signal peptide sequence"
/codon_start=1
/transl_table=11
/product="putative secreted protein"
/protein_id="CAC12793.1"
/db_xref="GI:10798657"
/translation="MTTLAVRLKNAFLILIAVLASFLICIRADGRYGVADYTYD
VHLPRGGLFTSHDVTYRGVSVRGVSPIDLTAEVVAELRIKSAPIPADKAVAG
LSAVGEQYIDLPRESGSEYLDGTRIDQADVEPAVPYDLSVPLASVPLELRA
TYVDEFGKAFDGDLDVLLDGSDFEADARALPSTLLINDETVLTQAOELRA
IRDPFVAGKDLAALAKGSPADRLRLAVPEATQVSGILRLDLSGLVLANLTT
EYAVVRGRIEELLYKPAVAAGATVADGCTLDIANTFSPSLPCDVGCTGRYN
GIDLGTAPLNTDACTAPASGCKVYKRGSAANPKGAVDPATPSSLPBGSGRTTPAG
GSPARDGSTALPAGALPLPGSGEAPAGLTGLAPAGAGAR"
complement(963..1397)
/gene="SC8A2.02c"
/note="Pfam match to entry PF02470 mce, score 98.10,
E-value 1.8e-25"
complement(1421..2662)
/gene="SC8A2.03c"
complement(1421..2662)
/gene="SC8A2.03c"
/note="SC8A2.03c, possible secreted protein, len: 413 aa;
similar to TR:053971 (EMBL:AL022073) Mycobacterium
tuberculosis hypothetical 40.2 kDa protein MT051.08, 377
aa; fasta scores: opt: 685 z-score: 681.0 E(): 2.1e-30;
36.0% identity in 311 aa overlap. Contains Pfam match to
entry PF02470 mce. Also contains possible N-terminal
region signal peptide sequence"
/codon_start=1
/transl_table=11
/product="putative secreted protein"
/protein_id="CAC12794.1"
/db_xref="GI:10798658"
/translation="MKRRATLPGRVAGLTAGSLAAVGLALTLIGSVVPSGFDIE
DLPFGADLGHPIYVTVAEIDVSLVPSHVAVRVDVAVGRITGELGEDDSAVT
MEINGEVRLPADATARLEOSSILGEYVYVLPAPKKTGGRLDGSVIPLATSRTE
VEEFGALSLNLGGGVNDLKTTRFLNALAGREVERSMLEKRVTLVGDLDDHNGD
ITDADLAVRLSLTLATRKDDVGTVLTDSLPGTLERKGSLLTMLRSLDLSGVAV
SITNAKSDMIADILAKVAPTLRALADAGDLPDLSVLTLPFTDVLGVAGDYLNT
YLSMAVPGTEVIVPLVDGDTGPRTASATADQDPDAAKNPAKNPISKRSSGGRGS
ASPLPLPSVSGAPASGEGHG"
complement(1431..1438)
/gene="SC8A2.03c"
complement(2138..2596)
/gene="SC8A2.03c"

		gene	/note="Pfam match to entry PF02470 mce, score 119.10, E-value 8.5e-32"
		gene	complement(2659, .3672)
		gene	/gene="SC8A2.04c"
		CDS	complement(2659, .3672)
		gene	/gene="SC8A2.04c"
		gene	/note="SC8A2.04c, possible secreted protein, len: 337 aa; similar to TR:053970 (EMBL:AL022073) Mycobacterium tuberculosis hypothetical 44.4 kDa protein MTW051.07, 423 aa; fasta scores: opt: 596 z-score: 625.4 E(): 2.6e-27; 33.1% identity in 320 aa overlap. Contains Pfam match to entry PF02470 mce. Also contains possible N-terminal region signal peptide sequence"
		gene	/codon_start=1
		gene	/transl_table=11
		gene	/product="putative secreted protein"
		gene	/protein_id="CAC12795.1"
		gene	/db_xref="GI:10798659"
		gene	/translation="MTRRKILTVLALVLAAGLAARALPGSTRTATYFPAIRAIGVYSGDRLILGVGRVEVESVDEGTRVRVGLDLDGIRKPEPARVAVAPSVADIRRYQTPVARTGPALADAVLPASHNHYPEVDIYDSITELGADLPDGAAGALSELISLARGALDNGNGEALIGDVGQEKAKTKIDSGDLEFLLSLQPTTYMLKKKIDYRAGERLDEVYSPFADKKDDLTGALAEGLGALGVKFTFDNKEKLKNDVRLPITRITVQKRSLEALDVAFLADNVNAYNPITRLIDGRANLNEISSGGPILLPLPVAGTEKDGSR"
		RBS	complement(2667, .2671)
		misc_feature	/gene="SC8A2.04c"
		misc_feature	complement(3238, .3654)
		misc_feature	/gene="SC8A2.04c"
		misc_feature	/note="Pfam match to entry PF02470 mce, score 103.70, E-value 3.6e-27"
		gene	complement(3669, .4724)
		gene	/gene="SC8A2.05c"
		CDS	complement(3669, .4724)
		gene	/gene="SC8A2.05c"
		gene	/note="SC8A2.05c, possible secreted protein, len: 351 aa; similar to TR:053969 (EMBL:AL022073) Mycobacterium tuberculosis hypothetical 43.7 kDa protein MTW051.06, 410 aa; fasta scores: opt: 683 z-score: 769.5 E(): 0; 33.4% identity in 368 aa overlap. Contains Pfam match to entry PF02470 mce. Also contains possible N-terminal region signal peptide sequence"
		gene	/codon_start=1
		gene	/transl_table=11
		gene	/product="putative secreted protein"
		gene	/protein_id="CAC12796.1"
		gene	/db_xref="GI:10798660"
		gene	/translation="MKRRGRKPLFRPKYERNPVAAGLLVTLVALLVNVNRLPEFGGTYSDAFSESAGLDGEDVRIRAGVYGQVTVGALDGAVKVSEVEDAWIDRTTGAIAIKTVIGKYLALDPLSGRDPGARILARTSPSTVADLQFQDLSPTVDLNNKRLAESEFETISDFPKDSPPHVKAAAGLSDLSKLSKPAKLSLTKSARTPTLIDNGKSSPTLIEDGPLGLGLERDRRIANLILKKSQDLCITBLGLVDKNEQLDPTLKAALGKVTSYLLEKNNRTRLEGTLLAVGPYIRKLVGNTLNGRMFDSYLGCVPRDYLPEIQTSPSTGCLPRKQPAQAQSGAR"
		RBS	complement(3682, .3685)
		gene	/gene="SC8A2.05c"
		misc_feature	complement(4254, .4670)
		misc_feature	/gene="SC8A2.05c"
		gene	/note="Pfam match to entry PF02470 mce, score 112.80, E-value 6.7e-30"
		gene	complement(4721, .5785)
		gene	/gene="SC8A2.06c"
		Query Match	5.1%: Score 52.6; DB 1; Length 7459;
		Best Local Similarity	46.7%: Pred. No. 5.5; Length 7459;
		Matches 237; Conservative	0; Mismatches 264; Indels 6; Gaps 2;
QY	286	gccgttcaggacgggcctcaagaagcttcgtgaactctccacggagatgccgcgtccac	345
DB	983	gacgacacttgcgcagactccctgcgcctgcgacgacccgcgcgcgcgcgcgcgcgc	924
QY	346	atcccttaagggaacaccggagagctcgtggaggaaggttggggagacggttcctcatg	405

Db	923	AAGGCTTGACGAGACAGCGGACGACCTTCAGGTCCTCTCGACAGCGGACGACATTTC	864
Qy	406	ctgagatgttggccctcaaccctggyggttcaaccgycagacgltccacgctggaagaag	465
Db	863	gtgagagccggccgacacgcccctgcgcgttcacacacgctctatcAACGACGCGGAGAC	804
Qy	466	gtccttgagaaaaagggcctgtgtgscacagacgtcttcacaaacgltcaagggag	525
Db	803	gtcctcgcgacacccagagcggcgccgacatcgcgacactgcgcgtccgaggaag	744
Qy	526	cgcgcggccatgagaccccttgggcgcgtccgctgtaggcgcagggaaagccagctcacc	585
Db	743	gaccttgccggccgctctcAACAGGgcttccGACGCGGACCTTCGCGCTCTCGGCTCAC	684
Qy	586	ctgagagatatactatacccttgaggaacgt---cgccatagacatgycacagcgctg	642
Db	683	ccgagagccggccacccagaggtacgscgctgtgcgggactctgacccgagcctcggggtc	624
Qy	643	ctctccttaacttggttcaagcctacaggaacacaggaatcgcgccacccctgagctg	702
Db	623	gttcctgcgacacctcctgACCAAGTCCGAGTGCCTGTCACCCGCGACGCGGATCGAG	564
Qy	703	ctggtcctcttggtctcaggggaaaggtgtagtgcacaaaccaaagacgctgacgtt--	760
Db	563	GAACCTCTGTGTGAATACCCGGGCGCGTCTCCGCGGTGTCCACCGCGCTGTGACGGGGAC	504
Qy	761	-acctggcctacatcctgtgtctccgc 786	
Db	503	AAGCTGACCTCGCGCTGCGCGTACAC 477	
RESULT 12			
SCM1/c	SCM1	34983 bp	DNA linear BCT 01-DEC-1999
LOCUS	Streptomyces coelicolor cosmid M11.		
DEFINITION	AL133278		
ACCESSION	AL133278.1	GI:6522830	
VERSION			
KEYWORDS	adenylyltransferase; alpha-mannosidase; dehydratase; dehydrogenase/reductase; glucose-1-phosphate adenylyltransferase; glycosyl transferase; hydrolase; integral membrane protein; lac-family transcriptional regulator; lysR-family transcriptional regulator; oxidoreductase; reductase; secreted protein; solute-binding protein; transport system permease protein; transposase remnant.		
SOURCE	Streptomyces coelicolor A3(2).		
ORGANISM	Streptomyces coelicolor A3(2)		
REFERENCE	Bacteria: Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.		
AUTHORS	1 (bases 1 to 34983) Redenbach, M., Kieser, H.M., Denapate, D., Eichner, A., Cullum, J., Khashabi, H. and Hopwood, D.A.		
TITLE	A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome		
JOURNAL	Mol. Microbiol. 21 (1), 77-96 (1996)		
REFERENCE	97000351		
AUTHORS	2 (bases 1 to 34983) Oliver, K. and Harris, D.		
JOURNAL	Unpublished		
REFERENCE	3 (bases 1 to 34983) Cerdano, A.M., Parkhill, J., Barrell, B.G. and Randsdrem, M.A.		
TITLE	Direct submission		
JOURNAL	Submitted (30-NOV-1999) Streptomyces coelicolor sequencing project, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof. David A. Hopwood, [3] John Innes Centre, Norwich Research Park, Colney, Norwich, Norfolk NR4 7UH, UK		
COMMENT	Notes: Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC and Beowulf Genomics Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web. (URL: http://www.sanger.ac.uk/Projects/S_coelicolor/) CDS are numbered using the following system eg SC787.01c. SC (S.		

coellicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand).

The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous.

The length in codons is given for each CDS.

Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon.

Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot program of Bibb et al., Gene 30:157-66(1984) as implemented at <http://www.nh.gov.jp/jun/cgi-bin/frameplot.pl>.

CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or (att)) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.

Cosmid M11.

FEATURES

source

1. 34983
/organism="Streptomyces coelicolor A3(2)"

/db_xref="taxon:100226"

misc_feature

1. 101
/note="nominal overlap with Streptomyces collicolor cosmid Stm10"

CDS

complement(1..1764)

/partial

/gene="SCM11.01c", hypothetical protein, len: 587 aa;

/note="SCM11.01c, hypothetical protein, len: 587 aa; similar to various hypothetical proteins, e.g. TR:092B0 (EMBL:AL035161) Streptomyces coelicolor conserved hypothetical protein SC9C7.20 860 aa; fasta scores: opt: 588 z-score: 654.5 E(): 4.4e-29; 30.4% identity in 542 aa overlap"

/codon_start=1
/transl_table=1
/product="hypothetical protein SCM11.01c"

/protein_id="CAB61912.1"

/db_xref="GI:6522831"

/translation="MSDRGASAPSLPDDMPAHPPDITALNMGSEFMDLADGLFHMDA QAHVFDLRPEEYDGRPESTALRVPTAESRMDITIVARMKDSENGTYPRLRRDG TLRWTHQGYTRDETRGPRRIIGIVRATQEMADIASRQALDEARRRLTVVOL ATALAHARIVDDVIDVLRDTHGLRGATSLVAGLVAAGRIRLVADDPESFPGTR VTRIDEPYPMSEAVRTLSPRFIESPEEFAEYPCGLMDITHLDTAAAYVLIQOAR IGATIGLYSDRHSRSPDRNVLVALGSSIASLORAMLYEEMDLAEGLOAMLPRTI PSVPGCVAYVYRAASITGALGRDIDGMDLPLPGRGVAVIGDVGHDTAAAVM GQULIVLAAVAEGHPATVVARASVFLHELDTRFALAEADLGVOVYVAGH IDPLLRGDSGTCRLRVYEGGLGLISAEFGTAYPAVATLELDPGNTLLCTGDVEOP GADLDGMDVLTALITSGPDVRLADRLIDVDERDDVALVLRHGLGAPRFE GRVOGHVSPGPEGLTEARHMI"

complement(1..1764)

/gene="SCM11.01c"

1874..1879

1874..1879

1874..1879

1874..1879

1874..1879

1874..1879

1874..1879

1874..1879

1874..1879

1874..1879

1874..1879

1874..1879

1874..1879

1874..1879

1874..1879

gene

CDS

complement(3319..6342)

/gene="SCM11.03c"

/note="SCM11.03c, possible alpha-mannosidase, len: 1007 aa; similar to TR:AA00190 (EMBL:AF04414) Homo sapiens alpha-mannosidase 6A8B, 1062 aa; fasta scores: opt: 1326 z-score: 1446.5 E(): 0; 37.8% identity in 1013 aa overlap and to SW:MAN1.RAT (EMBL:M57547) Rattus norvegicus alpha-mannosidase (EC 3.2.1.24), 1040 aa; fasta scores: opt: 1285 z-score: 1401.8 E(): 0; 35.9% identity in 1041 aa overlap"

/codon_start=1

/transl_table=1

/product="putative alpha-mannosidase"

/protein_id="CAB61914.1"

/db_xref="GI:6522833"

/translation="MHDRSLVEARLRKVLDERVVPALVPESVPLDVAVMNAPGPVP VEEGLAARPRPVEVARGAPWGTSTVETVTEVEMGKVEAIIIDGEPENPGTQ CEGLYVRPDGTPTVKGLNPNQWVRIGAVEGEGVRLVLEASNPVILDPYPTRL GDKDTAGSEPOYTLRMDLAVLDETVNMLVLDVLGELTALMLPVDSPRRETLRAVD KALDAIDLDVNGTAAEQARSRLTEVLAAPASVRIASVHAHIDSMILPLEYTK KVAETSMKATLIDEDPVPYAKSOQAMVVRPHRVPYKAKAVADGFTVAGG WVESDTNMGSEBAMROFVHGKRFLLDFEGVENDAMVLPDVFAGLIPDITKAAGK YLTKQISWSQTNKPHHTFMEGIDGRTIFHPVDTYVNCSSKMGSEIAAANFND GVARHSLAPMGMDGGGTTREVMVAKARLNLGSEVYEMEPHAFEEKAEANENP PVMVGELEYLHLRATLTSQAKTKOGNRSRHLRLEALATAAATVARGFYPDDLD RIMKTVLLHOFHDLIPGSSIAVREARATVRYAALNLGIIIAOARLAGEGPTPY FNSAPHARAGVPAGAAASPAIDGTGSLSPRGGVHVDNGLLRTEIDRGCIYSATLD AADRETTAPGAGMLDLPDPFMMMDAMDVDERVNTVYDLDADDEVDPDDASVR ITRSGSSRVTVQVLTLLAGERRLVDVDEVMHEKEPLKLAFLPDVAERVAESTOG HENRPTHTNLSWEAKFEACNHRFVHLDEEPMGVALVNDSTYGHDTVTVTDAGT TTVVRLSLIRAPRDEPDETDOGVHRFRHALVPGAGIDVAGEGRINLPEHLLTGAP EVAPLVYDRVAVVYTVKLAIDSGDVYVFEHAGGRTATLTAGFAVADVQVTLL LERPLADEARPPGDRITAVSLRFPOLTLAKKA"

complement(6350..6353)

complement(6484..7731)

/gene="SCM11.04c"

complement(6484..7731)

/gene="SCM11.04c"

complement(6484..7731)

/note="SCM11.04c, hypothetical protein, len: 415 aa; similar to TR:CAB5687 (EMBL:AL121596) Streptomyces coelicolor hypothetical 71.4 kD protein SC9A.35, 644 aa; fasta scores: opt: 205 z-score: 234.7 E(): 1.1e-05; 26.0% identity in 408 aa overlap"

/codon_start=1

/transl_table=1

/product="hypothetical protein SCM11.04c"

/protein_id="CAB61915.1"

/db_xref="GI:6522834"

/translation="MPSAVRFGVYTPVSGWFHMLDFDLDSVRADDSIAALDVHY RVPLMPYFQPNRLIRERAVEDVAVLVDAGEGDLVNVADGLOGHSFSDYVAPTR TWHRRNLFTPDVYVGOAATLRTAALAGVAPNLTGWTIGENVQSFAGHPDPDRAT SAOIDAMLERNLACCEGAGRGMLIAEVATYVODODOPPTPAOAGCVATYVHSY FNGRAOCHGRISVSEHNAAYLITLSAAMDDBRRPVMDEVGAPFALVFAEHAATF EATVENALDCEPDLMGVWCSHDVSRALAPPELEGILITDRRKRDTARVLAASA RAARDGARPAAPVPTTALAVPADPAARSACAPGAVYDAFERRIVADGARPTVLTGPA

z-score: 149.8 E(): 0.58; 27.7% identity in 166 aa overlap"

/codon_start=1

/transl_table=1

/product="putative integral membrane protein"

/protein_id="CAB61913.1"

/db_xref="GI:6522832"

/translation="MEMPTADPWTALRYLFOFALAVYLVAFETRALDPRALLGERGL TPVPRFVERPFRARRASLPQMTSDRLRAGCMAGCAVASALAGLDLPLMGAMLL WLVPALSLSTVNGQVWYFSGWESLLETGAVAVFGENEVAEPVYVLFRLPFLFR VEEGALILKMRDECEMRKTLCTDHHETQMPGPLSPFNHLPKRLHVEVAANVTQ LVPPLILPAPHPVSTPAALAMATOLMTVLGSEFNLMMWVITLVNYSVRPADPSV AAAPLMEVYVAVVALVALVPLVSHRPMVMSIRROVMNSPSTHLYNYSVAGVSRV RVEVYICTADEVAKODDMREYFEKRGCPRRRPPRPATYHRLDMLMPFALSPS VAGSFRGFVERLENDPATTILKLSRSPFPDAPPRVARRFLRYTWTBRLBTGA CWERYVREYLPRLTAPR"

complement(3319..6342)

/gene="SCM11.03c"

complement(3319..6342)

/gene="SCM11.03c"

complement(3319..6342)

/note="SCM11.03c, possible alpha-mannosidase, len: 1007 aa; similar to TR:AA00190 (EMBL:AF04414) Homo sapiens alpha-mannosidase 6A8B, 1062 aa; fasta scores: opt: 1326 z-score: 1446.5 E(): 0; 37.8% identity in 1013 aa overlap and to SW:MAN1.RAT (EMBL:M57547) Rattus norvegicus alpha-mannosidase (EC 3.2.1.24), 1040 aa; fasta scores: opt: 1285 z-score: 1401.8 E(): 0; 35.9% identity in 1041 aa overlap"

/codon_start=1

/transl_table=1

/product="putative alpha-mannosidase"

/protein_id="CAB61914.1"

/db_xref="GI:6522833"

/translation="MHDRSLVEARLRKVLDERVVPALVPESVPLDVAVMNAPGPVP VEEGLAARPRPVEVARGAPWGTSTVETVTEVEMGKVEAIIIDGEPENPGTQ CEGLYVRPDGTPTVKGLNPNQWVRIGAVEGEGVRLVLEASNPVILDPYPTRL GDKDTAGSEPOYTLRMDLAVLDETVNMLVLDVLGELTALMLPVDSPRRETLRAVD KALDAIDLDVNGTAAEQARSRLTEVLAAPASVRIASVHAHIDSMILPLEYTK KVAETSMKATLIDEDPVPYAKSOQAMVVRPHRVPYKAKAVADGFTVAGG WVESDTNMGSEBAMROFVHGKRFLLDFEGVENDAMVLPDVFAGLIPDITKAAGK YLTKQISWSQTNKPHHTFMEGIDGRTIFHPVDTYVNCSSKMGSEIAAANFND GVARHSLAPMGMDGGGTTREVMVAKARLNLGSEVYEMEPHAFEEKAEANENP PVMVGELEYLHLRATLTSQAKTKOGNRSRHLRLEALATAAATVARGFYPDDLD RIMKTVLLHOFHDLIPGSSIAVREARATVRYAALNLGIIIAOARLAGEGPTPY FNSAPHARAGVPAGAAASPAIDGTGSLSPRGGVHVDNGLLRTEIDRGCIYSATLD AADRETTAPGAGMLDLPDPFMMMDAMDVDERVNTVYDLDADDEVDPDDASVR ITRSGSSRVTVQVLTLLAGERRLVDVDEVMHEKEPLKLAFLPDVAERVAESTOG HENRPTHTNLSWEAKFEACNHRFVHLDEEPMGVALVNDSTYGHDTVTVTDAGT TTVVRLSLIRAPRDEPDETDOGVHRFRHALVPGAGIDVAGEGRINLPEHLLTGAP EVAPLVYDRVAVVYTVKLAIDSGDVYVFEHAGGRTATLTAGFAVADVQVTLL LERPLADEARPPGDRITAVSLRFPOLTLAKKA"

complement(6350..6353)

complement(6484..7731)

/gene="SCM11.04c"

complement(6484..7731)

/gene="SCM11.04c"

complement(6484..7731)

/note="SCM11.04c, hypothetical protein, len: 415 aa; similar to TR:CAB5687 (EMBL:AL121596) Streptomyces coelicolor hypothetical 71.4 kD protein SC9A.35, 644 aa; fasta scores: opt: 205 z-score: 234.7 E(): 1.1e-05; 26.0% identity in 408 aa overlap"

/codon_start=1

/transl_table=1

/product="hypothetical protein SCM11.04c"

/protein_id="CAB61915.1"

/db_xref="GI:6522834"

/translation="MPSAVRFGVYTPVSGWFHMLDFDLDSVRADDSIAALDVHY RVPLMPYFQPNRLIRERAVEDVAVLVDAGEGDLVNVADGLOGHSFSDYVAPTR TWHRRNLFTPDVYVGOAATLRTAALAGVAPNLTGWTIGENVQSFAGHPDPDRAT SAOIDAMLERNLACCEGAGRGMLIAEVATYVODODOPPTPAOAGCVATYVHSY FNGRAOCHGRISVSEHNAAYLITLSAAMDDBRRPVMDEVGAPFALVFAEHAATF EATVENALDCEPDLMGVWCSHDVSRALAPPELEGILITDRRKRDTARVLAASA RAARDGARPAAPVPTTALAVPADPAARSACAPGAVYDAFERRIVADGARPTVLTGPA

	TITLE	JOURNAL	COMMENT
gene	Direct Submission Submitted (02-VUG-1999) Streptomyces coelicolor sequencing project, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CA10 ISA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof. David A. Hopwood, [3] John Innes Centre, Norwich Research Park, Colney, Norwich, Norfolk NR4 7UH, UK		
CDS	Notes: Streptomycs coelicolor sequencing at The Sanger Centre is funded by the BSRG and Beckwulf Genomics Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web. (URL: http://www.sanger.ac.uk/Projects/S-coelicolor/) CDS are numbered using the following system eg SC7B7.01c, SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand). The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons is given for each CDS. Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon. Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the Frameplot program of Bibb et al., Gene 30:157-66(1984) as implemented at http://www.nih.go.jp/ Jun/cgi-bin/frameplot.pl. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or tct) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.		
misc_feature	IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid 4A10 lies between and overlaps with cosmids 4G6 and 2C1 on the Aser-C genomic restriction fragment.		
source	Location/Qualifiers 1..43147 /organism="Streptomyces coelicolor A3(2)" /strain="A3(2)" /db_xref="taxon:100226" /clone="cosmid 4A10"		
gene	1..102 /gene="SCA410..01" /note="Overlap with Streptomyces coelicolor cosmid 4G6"		
CDS	1..205 /gene="SCA410..01" <1..205 /gene="SCA410..01", hypothetical protein, partial CDS, len: >67 aa; unknown function, probable CDS suggested by GC frameplot and positional base preference. Continues in cosmid 4G6 as SCA4G6..39"		
gene	/codon_start=2 /transl_table=11 /label="SCA410..01 /product="hypothetical protein" /protein_id="CAB51977.1" /db_xref="GI:5689940"		
CDS	/translation="TRANRRVRVWDTRHGYGLDTADRAQMDYVVSDDRPAPRT SEWTKSTRSGTGVERVYSFA" complement(190..780) /gene="SCA410..02c" complement(190..780) /gene="SCA410..02c" /note="SCA410..02c, hypothetical protein, len: 196 aa; unknown function, probable CDS suggested by GC frameplot, positional base preference and amino acid composition. Annotated in cosmid 4G6 as SCA4G6..38c"		
CDS	/codon_start=1		

```
/transl_table=11
/label=SC4A10.02c
/product="hypothetical protein"
/protein_id="CAB51978.1"
/db_xref="GI:5689941"
/translation="MNTVLFHSTLGPAPVQAADRLNADGQVWTPDLFEGREDTY
EAGNADQDGLRELLRAVLAAPYSERGLVYVSGASIAQTLADGRRAGLL
LHGSLDAPNANSDVDELPIVLAEPDPETDDMWLSAWLQMGRIQADVEVHRYGAGH
LYMPDGLDDYAEAEAEATWRAVLFLESLOEAQGE"
gene
886..2034
/gene="SC4A10.03"
886..2034
/gene="SC4A10.03"
/unknown_function, shows weak similarity to parts of many
hypothetical proteins e.g. SW:TB43.METUA (EMBL:067556)
Methanococcus jannaschii hypothetical protein (361 aa),
fasta scores; opt: 222 z-score: 255.3 E(): 6.7e-07, 24.8%
identity in 306 aa overlap. Contains hydrophobic, possible
membrane-spanning regions. Contains pfam match to entry
PF00924 UPF0003, uncharacterized protein family UPF0003"
/codon_start=1
/transl_table=11
/label=SC4A10.03
/product="putative membrane protein"
/protein_id="CAB51979.1"
/db_xref="GI:5689942"
/translation="METILRLPIVYGGSVLTIVYIGMTDLILKRADGRHETPLMGL
LRARVPQIVLSCGLRGSYVEAGVPEHRTGRTLTLLIGSAVLVRIAAV
ETSYRVAHAHERDPAVRVRYTOVSLIMRVAVAVGVVSSMLTFEPARAAS
LNASGIIIGVAGVAAOSTLGNLFAFGQIAGDMVMDQVVDGEMTVEITLTYL
SVRTDERITMPVAYSTKPEENSRGCTPQMTGTFVWHLDSAPMDLRELRILR
ECPANDGRNVLVTDDPTTMEVRAIYATADADIMVRYVRYREGSMRLADEHPYA
LPRVNTADALRPPRAGLPHARHAPPAAGSSSDGARFRFERPAKSL"
1135..1881
/gene="SC4A10.03"
/unknown_function, shows weak similarity to the N-terminal portion of
eukaryotic Na+/H+ antiporters e.g. SW:MAH3_RAT
(EMBL:M85300) Rattus norvegicus sodium/hydrogen exchanger
3 (831 aa), fasta scores; opt: 271 z-score: 293.1 E():
5.3e-09, 25.8% identity in 481 aa overlap. Similar to many
putative Na+/H+ antiporters e.g. SW:YU23.MYCTU
(EMBL:277163) Mycobacterium tuberculosis putative
Na+/H+ exchanger (542 aa) (31.4% identity in 528 aa
overlap). Shows weak similarity to SC66T3.14c
(EMBL:AL079348) S.coelicolor probable Na+/H+
antiporter (514 aa) (27.5% identity in 545 aa overlap).
Contains hydrophobic, possible membrane-spanning regions.
Contains pfam match to entry PF00999 Na_H_Exchange,
Sodium/hydrogen exchanger family"
/codon_start=1
/transl_table=11
/label=SC4A10.04c
/product="putative Na+/H+ antiporter"
/protein_id="CAB51980.1"
/db_xref="GI:5689943"
/translation="MDQLALFMLLLGALLSPICARGLPAPVMTLLGLYALIDE
VNVNIPPELILPGLPLVAVARTSWRQPAANKRPIFLAVALVEVTVNCVASVA
HAIVGELPIAAAFALGALVAPDPVAATVAGQLGPRRLVILBEGSLPNDVTAIVL
YVAIAAAVAGSGSPRAGFDLVLSAVVAIGVIGWSNKLMSVIGPFIQIGITL
IVPVSYVADELHSGVLAIVLTAMFLAEVSDDDVMTRLGSGHTMVVDDVLTAVG
AFGLIGLELHNAVPTASGRMTELLGMAAIIIVVYVRLIMLIPMTLQRLHARDQ
DEDIPTSRETYVMWMAKRGVAVSALALATLELDDGAFFHRETYIAGVIMGT
LLQGLTLPWLVGRUGVRAVDTREQEEFERDLALRAKAKQRKLETSVEELPEELSE
QMIRRAFDIGIRISDPLGDERREAOEQRRARIKRVRIQGEMLSAARHEVLAARSEP
```

```
misc_feature
complement(2410..3633)
/gene="SC4A10.04c"
/unknown_function, shows weak similarity to Na_H_Exchange,
Sodium/hydrogen exchanger family, score 258.90, E-value
6.7e-74"
complement(3731..4213)
/gene="SC4A10.05c"
complement(3731..4213)
/gene="SC4A10.05c"
/unknown_function, shows weak similarity to hypothetical
proteins from other eubacteria e.g. SW:YVBD_BACSU
(EMBL:D26185) Bacillus subtilis hypothetical protein (147
aa), fasta scores; opt: 199 z-score: 236.4 E(): 7.6e-06,
36.8% identity in 144 aa overlap. Shows weak similarity to
SC5F7.12c (EMBL:AL096872) S.coelicolor possible
acetyltransferase (169 aa) (37.7% identity in 114 aa
overlap). Contains pfam match to entry PF00583
Acetyltransf, Acetyltransferase (GNAT) family"
/codon_start=1
/transl_table=11
/label=SC4A10.05c
/product="hypothetical protein"
/protein_id="CAB51981.1"
/db_xref="GI:5689944"
/translation="MPSRAPAAVEYRVAEDPVDRACFAVRKDVFAVBOKVPEDIEYD

Query Match 5.1%; Score 52.4; DB 1; Length 43147;
Best Local Similarity 44.5%; Pred. No. 4.6;
Matches 209; Conservative 0; Mismatches 261; Indels 0; Gaps 0;

QY 387 gttcgtctcatagtgctccctctgagatgtytggccccaacccgagggatcccgagac 446
DB 42648 GGGCGACCTGCGCCGAGTCTCTGGACCCGCTGCTCGCGCGCGCTGACATGTCGACGT 42589
QY 447 cgtccacgcttgaaagaagtcctctgagaanaaaggcctgtgtgcccacgagctctca 506
DB 42588 GCGGGAACAGGGAATGGAGCGCGCGGAGAACTGGAGCACTGCGCGCTGTCGCCGACGC 42529
QY 507 ccaaacgctcaacgagggagcgccgagcgatcggaccccttggccgtccgtcgaagcc 566
DB 42528 CTGCGGCCCGACGAAACCTCTCCGCGCAACGACCGCGCGAGCTGCGCACCGCGCC 42469
QY 567 agggaaagcgaagctacccttgagagacatcatcctccctgaggaacctgcctaga 626
DB 42468 CCGCGCGGACGTCTCACCTGGCCAGGCGGACGCTCGGCTGCCGCCGCCGCGCAT 42409
QY 627 catgccaacgagctgctcctctcaacttggtcaagagctaccaggaaccaggaatccg 686
DB 42408 CCTGCGGCAAGACGTCTGTATGCGCGCGCTGCCAGCAGCGCGGATCCGAGGCGCGCGC 42349
QY 687 cccaccctggagctgctgctcctcgtgagctcaaggagaaagggtatgcccacaacca 746
DB 42348 CCGCGGCCAGAGAGGTGTGACTTCTGCACCGCGCGCTGTCGCGCACCCCGACCAA 42289
QY 747 gaccgtggccgttgacctgagctcatcctgtgctcccccgaagtgagagttccaaact 806
DB 42288 GCCCGGCCGCGACGACCCGCGCTTGCACCTGTGTCGCCGCGCGCGCGCTGGGTACCA 42229
QY 807 cccggccctatcacctcatgtctacgtacatgtgcgactcctagatg 856
DB 42228 CCGCCCTGTGTTGCGATCGCGCGGATCGGCACTCGGCAACTCGAGAGG 42179

RESULT 14
AF357202 113193 bp DNA linear BCT 17-JUL-2001
LOCUS Streptomyces nodosus amphotericin biosynthetic gene cluster,
DEFINITION complete sequence.
ACCESSION AF357202
VERSION AF357202.1 GI:14794889
KEYWORDS
```

SOURCE Streptomyces nodosus.
ORGANISM Streptomyces nodosus
Bacteria: Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomyces; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 113193)
AUTHORS Caffrey, P., Lynch, S.V., Flood, E.M., Finan, S.M. and O'Leary, M.
TITLE The amphotericin biosynthetic gene cluster from Streptomyces nodosus
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 113193)
AUTHORS Caffrey, P., Lynch, S.V., Flood, E.M., Finan, S.M. and O'Leary, M.
TITLE Direct Submission
JOURNAL Submitted (07-MAR-2001) Industrial Microbiology, University College Dublin, Belfield, Dublin, Ireland
FEATURES
source 1. 113193
/organism="Streptomyces nodosus"
/db_xref="taxon:40318"
/complement(4. .1824)
/gene="amphg"
/complement(4. .1824)
/gene="amphg"
/note="possible amphotericin export protein"
/codon_start=-1
/transl_table=-1
/product="Amphg"
/protein_id="AAK73498.1"
/db_xref="GI:14794890"
/translation="MAPDVEHEHEERESROPYRLAAILRPHRSVGATAGVYGI
LWAFGRILLGRVTDLIAQGYLGHGPAFVYDGAQLRLMILLYVYASVMYLVN
WYASVYRLIHLRRAQKRLRLPRFDRKPAEETLSRGTDVDNQLQOTLT
DLISVSFLIMLSLMLIISPLAGVRLSVPSGLLAMI SKRAQPOYAAVMASAGK
LGVAVEMCGHALVKAFAERAEQERFDEANEAVYRAGAOFAAGALEPVMFAN
LGVAAYVAGAMKVNKSLIGDQAFIYARQSOPIVEIASVAGRLSVSPSTAVV
FTLLDPEOEPEPRPLAVERVEGRVPEOPVSRYPDPLIEGLSLVSPSTAVV
GPGACKTAYANLMREYEDSGRIIDGDTAMRDRLRSGVLYDPTPEKTI
AENIAGSREATRAQDYEAARATYADRFTITLSQIDYTDDESQVSGEKLITVA
RAFLARPAVYLDDEATSVDRTELLIQRAMNLRGRISFVIAHRLSTIRDAIVV
MKSRIVEQGTQDLIDAGAVARLHAADAPADVTG"
complement(1805. .3628)
/gene="amphh"
/complement(1805. .3628)
/gene="amphh"
/note="possible amphotericin export protein"
/codon_start=-1
/transl_table=-1
/product="Amphh"
/protein_id="AAK73499.1"
/db_xref="GI:14794891"
/translation="MAPSVRWTPPPFOAAAGRQPCGSVLRIMTQLRPYASVYA
LVYLVQLIGLILPLTGLAALIDBGVVRHDSRIGTIGATMAVVALVOIAALGAA
LGARTSTALGRDRSAVFRVRDPSAREIGRETPSLTRSVNDVOQONLAQSGGI
FVAAPLMCGSVLALRODYTLALILVPMVLVAVCFGLLSRMALYARLQOTLRI
GGLRLRITGVYVRSFARDAGEERTTNEELGLSGVGRILVAMPVSVLLMNL
FTLLGTVGARRIDSGMOIGALSAFLSYLSILMSVMLAFVFLVPPARVAKBIT
EYDAETDVYPPASPRPMAGPGOVELNGAEFRYPYPAENAVYLDLSTLRPGRAVL
GSGSGKTTLLILRLVDYTAGEVIRIGTDVRELDPSYMAAVGVPRPYEFACV
ASMKRGRPDADDELEWVRLAQRHGFYRLGDLTETLAQGGTIVSGGQRRLATAR
ALIRAPAILFDPSFALPOSTAALRKALVPTEGATVTVQAQVASYRADRIYLL
DOGGIATGTTHDLRDSPTREIALISQRTRETAAGRS"
3840. .4874
/gene="amphditi"
3840. .4874
/gene="amphditi"
/note="mycosamine biosynthesis protein"
/codon_start=-1
/transl_table=-1
/product="Amphditi"
/protein_id="AAK73500.1"
/db_xref="GI:14794892"
/translation="MPKRALITGITGQDSYLAENHLSQGYQVWGLIRGOANPRKFRV
SRLASLSEVDGLMDQGLSVAVKQVPEDEYNLGAISFVPMKMOQALVEYEVNMG
YLRVLEAIRWVSGLSMSTRAGTEGQIRFYQASSSEMFQVAETPQRETTLLPHRPSYG

AAKAYGHFTRNRYRESGMYAVSGMLFNHNSPRRQGEFTRKISLAVARIKGLQDLK
ALGNMDAVRDMGYAVRAMHMLQDDAPDDVITGEMHYRQDAVRAPAEVIGDM
KQYVYVDPDLVRAEVEVLCADSSKQAOLGMKPSVDPELRNMVYDADLASRQNE
LDLULHSHS"
5042. .33574
/gene="amphh"
5042. .33574
/gene="amphh"
/note="polyketide synthase multienzyme polypeptide housing
extension modules 9, 10, 11, 12, 13, and 14"
/codon_start=-1
/transl_table=-1
/product="Amphh"
/protein_id="AAK73501.1"
/db_xref="GI:14794893"
/translation="MDNEOKLRDYKLKATANDLRARRRGELESASQEPYATGMTCR
YFGVSSPEDLRMVEAGEGVTPEFTDQKMDLEALASSPTSGGFLHAPEDADDF
GISPREAVAMDPOQRYVLESAMEAFERAGINPTSVKSGRTGVEIGMADORYGPDG
AGEFQULGTNGSVLSGRISYTFGTGPAVTVYDACSLSLVAHLATQALRAGECTIAL
AGGVYVSGPTEIEMGRQGLSVDRCSFGTDADGTGMAEGVGLIYERLSDAIRN
GREILAVRGTAVNODGASNGLTPAPGPOQAVIEGALVNRLSAGDIDVEAHGCT
TGDDPYEAQALLATYGOORDEKPLLIGSVKSNISHTQAAAGYACGYIKMMVRHCVL
PRTLLADEPFRHDMQCAVRYLTENTEMPATGAPRAAVSSFGISGTAAHTIYEDAP
EPEPADPEDDAPSTPAVTVGLVPLLSGSPSEYLRQAALLTITGTPPADLAY
SLATPTAEHRAVLASDLPELTGRLTAIAETGDAVLADVTGTARETRAVLEFT
GQCAORLCAGRELAARFVFAALADALDALEPHLDVPLRKYLMGEDADRLDREYAO
PALFAVEVALYRLISEFEVKPDLHLAGSGETIAAAVAVCFSLDADATVLAARGRLM
ALPEGGAAYAVOASDEEVAPILAGHEDVLSIAVAVPSVYSGDPTTETLAARLAA
DCKRTSRVASFHSPILAMPMLDERNVEGLITLHSLPLVYSDTGTATYAOQITS
PDYVHVHVAQNAFADGITMLAHADVTALEPDSVLSAMONCLDAGSALATVPA
LREGREDHTFALHAAHLHOGTALHMDACFTGTGRRLDPLPYTFQRRYPRVAVOG
GADLSVGLIGAHNHLISAASIASDEGALLTGRISLSHPMLADHYATVLLPGT
AFLEILVRAEGVDCRVDDELTAAPLVPEQGVYOVLMGNPDGSGRSTYVGR
DDEDEPMSHATGVSASRTSPEDATVPADARETLVDDGLYERLAGEGCGYGLF
QGRAMRWGDEFAEVNIPESGHTDABESGRLPALDSALHAASVYDDDERAAGSLP
FSWEGVSLHAGSATTTLRRLAPAGDAVAIAAODSGQVLSDDLILSLARADLEDA
AALVYDLAFELDMVEVTVYVAAAGTAAVAGVDFELRALPOGDLAVHPLADLAA
DGAVPDTVLLPLTGPGPDADPVTAHRAATEALAAATWLEDEDERASRLALVIRGA
TGHDDPAAAVMGVLSAOSNGRELVLDLADODT PALPAALITSEPOCALVAGEE
LRARLVRRPASTAEAVPAFEGGAVLVGTGGGAVLVARHLVAEHGVEILVSR
GGAAAGAELVAELAEVAPATVYACDVDRRAVAVELVAHPSAVHSAAGVLDGMAV
GTTPEPRTTVYLRPKYDAMNMLHEARDLDKAFVLFSSVAVYLSRGCANYAAGNAF
LDLALAHRRRAAGLPGSLAMGPBEGTGGTGGISEDDLRKMAAGTALTVBQGLAL
DADLDDDDAALPVRDLISVLRAGQEVPLLSLJGRSRRRAVAVASAGLQALRLA
RLDABESRDELVDLYRGQVALYLGHTAGELIAGRAFRLEGDSLAVELRLNRLLVT
GLRLPATVLEVDYPTVSHLASVYLDLGEVAAEVQRTAAVADDP IYVMACQYPR
GGVTSPEDELMRLVTTBGTDAVSGFPVNRGDENLVDPDDHGTATYTRSGGLHEGE
PDGPFGRMSPREALATDQOORLLIEMSMARACDPYGLGSAATGACVGMXGYS
AMIGSPEFGFGSGSSPSLASGRVSYTIGLEPATTVTACSSLSVAMHMMOALRS
GETISLACGVTVMSTPAVFDVARQNGSLTAPNGRQCAFSDSADGVMSRGMVLDER
QSDAIRNGHQIILAVVNGSAVNODGASNGLTAPNGRQCAVIRQALASGLSAGDVVY
EAHGCTTTGLDEAQLLATYGRDPEQPLILGSVKENICHTQAAAGVAVIKVM
SMRKHVLPRTLHNDPASHVWMEGAVELTQOTAPLEGGRRRAVSSFGISGVNH
TVLEADRGTTVAPAPABERTAGVAPLLISGRTIRLRAQAAALTHLQNHPPSADL
GLISLARTREFFERRAVITQODREGLASISLAAGRPDAVYEGEAGAGARAVAFSG
QOSQORAMKRELIETOPRFAAFDEVCALDPLDLPEVLEVPFAEGSEBALDLBRTG
WQPALFAEVALRYLVESGVARADVFVTSIGELAAAHIAVFTLQDAARVAAART
LMEALFSGGAMAVAOATREBEVAPLLEGEISVAAVNPVSVVSGEDPVELEAASG
ROBRTKRLRVSAHFSPHMDAMLEARTVLEISVAAPIPIVPSYDLTGRADAEVRT
ADYVWRHVEAEAFBADCVRTLDACAFTLELLEGGPGLITAMAEVLETDGDRVHNHVAL
PILRADRPESAAATTAARLQIHGVDDMTATYLAGCARVDLPYTAHQAHYWPOLP
SAAPSPADGPADOKLMAAVERGDPAELAVNIDDESITPLDLSLPAUSMKRGNOK
ALIDTLRYRVENWTRLSKPAAPVLDGTMVLVSSDANADDETTLLDGLAELMGAHGRV
RLVLADACDRAVLGRLADVTENADNTAQLVSLPDEPRTGPGAGFTGGLATTLAY
QALADTGAAGRLMTATRGAVSTGPADPYVPOATANGRGVATALEHPRLMGGLADLP
ADEDRGAGRLAEVLAVRDAPDGEQVVALRAGVARGARVRIHTVDELPEADQFTASGS
VLTGTGGLGAETAARMLARSGAAHLVITSRGCGVAPCAEARLAELESGASVSYTAC
DYADRDLAAVLDGUSADOPLTGVYHTAGVGRGPDALITPREFAGTAKLAGAHL
DNLQDRELDFLIFGSLAGVSGQOASVYGAMAVLDLALARRARGAIAVSIAMGP
WGTGMAADAVASGTLRQGLGLWBPALVLEIMRAVYVQDVTYVADVDMRVAPLF
TVSARSHALSIDPEVVALAENTPADTGAOSTIVRVSLSPEQDLRLITLDVREFAA
TYGSSSACAVDEDAFREIGDSITLVAELRKLHAGLATSLSIPVMVGPYPLPLEAO
YLRAEWVSVLEVAGVATGCTDDEIATIGNSCRYPGVSPPBQMLMDLVLSGTAAIT

Oy	620	ccttagaatgccaacgagcgtgctctccactcactgggtcaagacccac	669
Db	15813	CCCCCAAGGCCCTCCCGACGACGCGTCATCCGCGACGCTGACCAAC	15862
RESULT	15		
AL646072		188050 bp	DNA
LOCUS			Linear BCT 07-DEC-2001
DEFINITION		Ralstonia solanacearum GM11000 chromosome, complete sequence;	
		segment 16/19.	
ACCESSION		AL646072	AL646052
VERSION		AL646072.1	GI:17429824
KEYWORDS			
SOURCE			
ORGANISM		Ralstonia solanacearum.	
		Ralstonia solanacearum	
		Bacteria; Proteobacteria; beta subdivision; Ralstonia group;	
		Ralstonia.	
REFERENCE		1 (bases 1 to 188050)	
AUTHORS		Salanoubat,M., Genin,S., Artiguenave,F., Gouzy,J., Mangenot,S.,	
		Ariat,M., Billault,A., Brottier,P., Camus,J.C., Catolico,L.,	
		Chandler,M., Chisne,N., Claudel-Renard,C., Cunnac,S., Demange,N.,	
		Gaspin,C., Lavie,M., Moisan,A., Robert,C., Sarin,W., Schlex,T.,	
		Siglier,P., Thieault,P., Whalen,M., Winker,P., Levy,M.,	
		Weissenbach,J. and Boucher,C.A.	
TITLE		Genome sequence of the plant pathogen Ralstonia solanacearum	
JOURNAL		Unpublished	
REFERENCE		2 (bases 1 to 188050)	
AUTHORS		Boucher,C.A.	
TITLE		Direct Submission	
		Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston	
		Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie	
		Moléculaire des Interactions Plantes-Microorganismes INRA-CNRS,	
		Bp27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean	
		Dausset-CEPH, 27 rue Juliette Dodu, 75010 Paris, France, ILMG CNRS	
		118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA	
		URVY, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France,	
		Laboratoire de Biometrie et Intelligence Artificielle INRA, Bp27,	
		F31326 Castanet-Tolosan Cedex, Laboratoire de Genetique Cellulaire	
		INRA, Bp27, F31326 Castanet-Tolosan Cedex	
COMMENT		Christian.Boucher@toulouse.inra.fr	
		http://sequence.toulouse.inra.fr/R.solanacearum.html.	
FEATURES		location/Qualifiers	
source		1..188050	
		/organism="Ralstonia solanacearum"	
		/strain="GM11000"	
		/db_xref="taxon:305"	
gene		73..909	
		/gene="Rsc2802"	

```
CDS      73..909
          /gene="RSC2802"
          /function="miscellaneous; hypothetical/global homology"
```

Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Homology

```
/codon_start=1
/transl_table=11
/evidence=not_experimental
```

```

/protein_id="CADI6509.1"
/db_xref="GI:17429825"
/translation="MTALLGSKASVCVGNSEFKTLPALGAAAGTVYRITGCAIILA

```

SRRALDEPMVIMGAVAGLLILATVGGRAGOIDELGAATLGAAGVCMALYIVYIGKRVSTL
PAGQATSLGMANGALFAPIPGVQAGPALLAPLIVGLGGLSSAVPLSELMVALR
HISGRTRFSVLLSLEPAIGALGAVLTHELHLSRQMAIVAITAASGCCAATARSRAA

```
gene complement(936:2024)
/ gene="RSC2803"
/ note="RS00302"
```

```

/gene="Rsc2803"
/function="miscellaneous: hypothetical/partial homology"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Framed"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product_id="CADI6510.1"
/protein_id="CADI6510.1"
/db_xref="GI:17429826"
/translation="MAIEISFEFQYDQARFLDARARGLAIEPRAVPHALGPAG
EPLIDPACLPDRGALVLTGSHGEGCGSCGVGLRDDALFARLAAAGVALL
LVHAVNYGFARLVNEDVNLNRNSADFAVASANPAIETVDLLPTTPPDAN
OALLOOYAHGEGALRDAMTICQYATPDPMFGGATCTSTACATILSRHADAR
LANIDHTTGAGHGEKITSGADPRLERATIGWADVRLIAAGSVSSVEALDF
RACALPFAVEKVTYLTLEFGTLEPMAVQALRADHWLHRHPTSPDQAAAIKGLDAF
YCDTPAKGKGVYQARVAVLQAAAFESG"
/complement(2124..3326)
/gene="nrdb"
/note="Rsc2804: RS00301"
/complement(2124..3326)
/gene="nrdb"
/ec_number="1.17.4.1"
/function="small molecule metabolism: central intermediary
metabolism: 2-prime-deoxyribonucleotide metabolism"
/note="Product confidence : putative
Gene name confidence : putative
predicted by Codon_usage
predicted by Framed"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="PUTATIVE TRANSMEMBRANE RIBONUCLEOSIDE REDUCTASE
(SMALL CHAIN) OXIDOREDUCTASE PROTEIN"
/protein_id="CADI6511.1"
/db_xref="GI:17429827"
/translation="MLSMDDVKAPOAPOPAYOPOPILOSPFPPDGGALPSAT
OAGGILGNPNNAASSRNVNADKRVNGATDVNOLPFRKRNMEKRIAGCANM
MPOEINMSRDIALMKDPNGTDEBRITIKRNLGFTYADSLANMVLGYRQTABE
CROQLKQAEFAIHTHAYOIVESLGLNAEILFNAYHEVOSIRKDEFLIPIDYLT
DPSEFKGTGPENDKLLKSLIFACIMEGLFYFTQILAMGRONKWTGAEOYQYL
RDSLHCNFEIDILNQIKLEPNHMTAEFAETTELKFAVLERYAEDTMPRGVYG
LNAPMEKGYLRFICNRCCQIGDALFPEDNPPMSEWIDLKKEKNFETPIVETQ
TGALSMW"
/complement(3477..6440)
/gene="nrda"
/note="Rsc2805: RS00300"
/complement(3477..6440)
/gene="nrda"
/ec_number="1.17.4.1"
/function="small molecule metabolism: central intermediary
metabolism: 2-prime-deoxyribonucleotide metabolism"
/note="Product confidence : putative
Gene name confidence : putative
predicted by Codon_usage
predicted by Framed"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="PUTATIVE RIBONUCLEOSIDE REDUCTASE 1 (LARGE
CHAIN) OXIDOREDUCTASE PROTEIN"
/protein_id="CADI6512.1"
/db_xref="GI:17429828"
/translation="MQTAQTEIHSSTANPTAFSOPAEGTAVASAGYADYKILRRNCA
VVGFEPSKIVAMTKAFIAVNGGAAASAPVLEBETGVNVRALLSRSPNGSTFHI
EVODQVELAMRSRGEHDVARVYLVREBSROAOGTQOOPAFGLVVTGGTIR
PLMALRLNLIYACAGIGESVDEPIKTKVKNLYEGVMSQVYSAIIAATLIEK
DPYQSATARIILHTIRRELTGEVTOAEAGRADYFPDFIRGINAGGLDPTLAF
DLAKLGAALDASRDLPFNLYGLQTLTYDRFLHTISDKRIEMPQAFEMVAVAGLNEID

```

```

REARATFYOULLSFDEMSPTPLFENGSTCRSOLSCYLTTSDDLDGIYEALKENAL
LSKPFAGIGDMTNVRALSGHIGTNRKSGGVYPLKLVNVDVAVAVNOGKRGAAYA
VLETHMWDIEEFLERKNTGDDRRRTIDMTANIPIPLFKRMVEGENTLFSPTCP
DLHKVKAFAEAVLTGEDVARGELKFRKPAQLQRMKGLMLETGHPWTFKDP
CNIRSPQOHVGVHSSNLTCTEITLNTNDSITAVCNLSVLAHVAVQAQGSYALDHD
KLKRTVTRAMRLNDVINDINYAVKKAARDNLRHPVGLGIMGFODALRIPYAD
AAVOADTSMCAVCYAYWASTELAEERGSYKSLMORGVLPODSLKLAQERGG
YLEADLSSTMDGRLRERIOHGMNRNVCVAIPATISINIGVSAIETPTONLYVK
SNLSGFTYVNDYLRDLKARGLMDEPVNADLYFPQSLARIRIRPDLDTATAFE
VEROMLYEASRRKKWIDQOSLNTIYAGSGKMLDITYLRAMIRGLKITYYLRITGA
THVEKSTVSRGTLANAVSSGSDASVAAAGAPVAPASALDTVAATAPAMPDAGAVCT
MRPDGPFCECECOQ"
/complement(6877..7479)
/gene="vampd"
/note="Rsc2806: RS00299"
/complement(6877..7479)
/gene="vampd"
/function="cell processes: protection responses;
detoxification"
/note="Product confidence : putative
predicted by Codon_usage
predicted by Framed"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="PUTATIVE AMIDASE PROTEIN"
/protein_id="CADI6513.1"
/db_xref="GI:17429829"
/translation="MEPVAGPVRIGADGMADGVCRRHSPVNDARPEGLPDLIVLNI
SLPPARGTDVLDLDEFANLDCGAPYEQLRGVSAHFRIETGACVAFPCADARA
MHAGVSDFEGRTRGNDPSTIGTIEGTDOPTPAQVAAATAVRAICAAVPIAIGH
SDIAPGRKTPDGCPCFDMHRLAKLEALFYPQAGAAPA"
/complement(7472..9184)
/gene="pentr"
/note="Rsc2807: RS00298"
/complement(7472..9184)
/gene="pentr"
/function="small molecule metabolism: global functions;
global regulatory functions"
/note="Product confidence : probable
predicted by Codon_usage
predicted by Framed"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="PROBABLE RESPONSE REGULATOR TRANSCRIPTION
REGULATOR PROTEIN"
/protein_id="CADI6514.1"
/db_xref="GI:17429830"
/translation="MSRPPFRPMPMSKAAVREPLIVNDEADLRRELTSLRMGHD
VYIAGLAEREAALTRORFALVLTDMKLVGLGSLGIDLYRQLSAARDPRPVAVITAYGA
ENAVEALKAGAFDYIAKPLSLDQIRLVYLAALGQODPDPSADRLAERTSALPLGHS
AAAOEVRSILRLRLARSMAPVYIGESGSGSARRARALHAISSSARFPVAVNCGALPE
NLMEAEFFGVKGAFTGADSDRGQFPAAGSTGLMDEVLDLPTMQVKLRLAQER
VRKIGESREDPVVRYVYASHONLARIVAAAGRPREDIYFLANTYLEMPLRLRAEDY
PVLAGVYLBOLARVYGLRPRKRLTROALGOLGAPFPGVAVRELDNLEERYAAREGS
IDVDHICALSTDIERSPLFRHAREAHGHPVYRANLPRVYARHPRAHPRGHPVAYPL
GVPOVGGWPDAAITVYVPKGLMPHPVYVYVEVERPAVYEPAPAPAPALPVDLPA
YLSVERGVTLAALQGTGFNRTASKLGLISFRQDLRYRMQOILIDRPRDIEAAANGG

```

Query Match 5.1k; Score 52.2; DB 1; Length 188050;

Best Local Similarity 47.8k; Pred. No. 4;

Matches 187; Conservative 0; Mismatches 198; Indels 6; Gaps 1;

```

QY 387 gtcgtcttcattggtcccccgtgagatgtgcccctcaacctgggggtcaccgcgcagac 446
      || || | || || | || || | || || | || || | || || | || || | || || |
Db 140037 GCGCGGCGATGCCG6GTCAACATGCGCACGACGCGCAGGCGCTACACCGATCTCAA 140096
      || || | || || | || || | || || | || || | || || | || || | || || |
QY 447 cgtccacgcctcgagaaagctcttgagaaagagcctgtgtgcaccgcagcgtccctca 506
      || || | || || | || || | || || | || || | || || | || || | || || |
Db 140097 CTTCTCATCCCGGAACTGTCACGCGCATCGCCTACAAAGAAAGCACCTTCTCGCAGA 140156
      || || | || || | || || | || || | || || | || || | || || | || || |

```

OY	507	ccaacgcgtcaagggaagagccgggacatcggacaccttggccgctcgggtatggcc	566
Db	140157	GGAGGCGACATCTTCTCGGCGCCGCGCGCTTCGGATGACTACGCCGACAAAGTTGCGGA	140216
OY	567	agggaagaagccaggtctcaacctgtgcagatacatlaacctgtgaggaacctgcgccataa	626
Db	140217	CGGCGTCGCCCAATTCGACAGCTGGGGCGCAATAGGCTAACCGCGCGCGCTGCTGCCAGATC	140276
OY	627	catggtccaaagcggtgtcttccttcaactgtgtgtcaaggtctaccagagcacaggaatccg	686
Db	140277	GGCGCGGCTCGGCGCGGCGACAGCT-----GCTGTACGGTTGTGAGGTGTGGCGCAGAA	140330
OY	687	ccccaacctgtgacgtgtctgtctctgtgtgtcaaggggaaaggtgatgtccacaaccaa	746
Db	140331	CGGCGCGTGGGTCGTGCGCGGAGGCGTGCAACAAGCTCAACGGCGTGCGGTACACCGT	140390
OY	747	gacggtgacgtgtgactgtggccatccg	777
Db	140391	CGCGTGGGCGGCGCGAGCGCGTCAACCGTg	140421

Search completed: May 29, 2002, 10:21:27
Job time: 7259 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 29, 2002, 09:10:08 ; Search time 210.56 Seconds

(without alignments)
8366.044 Million cell updates/sec

Title: US-09-664-186-4

Perfect score: 1026

Sequence: 1 gtgaagaacacaaacaccc.....ccctcaagagagcctcctga 1026

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

tal number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_032802.*
1: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1026	100.0	1026	22	AA04666
2	1026	100.0	1026	22	AA04668
3	54.4	5.3	615	14	AA037036
4	51.6	5.0	1815	22	AA009400
5	48.6	4.7	21034	19	AA062154
6	48.6	4.7	117213	19	AA062176
7	48.6	4.7	154746	24	AA025519
8	47	4.6	6094	21	AA071428
9	46.2	4.5	1473	19	AA034317

10	45.6	4.4	1620	13	AA022482
11	45.6	4.4	2668	13	AA022485
12	45.4	4.4	1686	23	AA054366
13	45.4	4.4	26338	19	AA062134
14	45.2	4.4	1290	21	AA043669
15	45	4.4	4767	21	AA071571
16	44.8	4.4	2458	21	AA093444
17	44.8	4.4	2838	23	AA054251
18	44.4	4.3	1833	23	AA022159
19	44	4.3	660	23	AA054238
20	44	4.3	4451	19	AA023288
21	43.6	4.2	415	21	AA047070
22	43.4	4.2	1314	21	AA064885
23	43.4	4.2	12588	15	AA063293
24	43.4	4.2	35026	21	AA064890
25	43	4.2	65140	22	AA017184
26	43	4.2	125401	22	AA017186
27	42.8	4.2	2614	11	AA006631
28	42.8	4.2	24379	18	AA093095
29	42.8	4.2	24379	19	AA025925
30	42.8	4.2	77536	21	AA014651
31	42.2	4.1	897	19	AA037167
32	42.2	4.1	897	22	AA0401758
33	42.2	4.1	1314	21	AA040194
34	42.2	4.1	2584	18	AA051330
35	42.2	4.1	2584	23	AA092733
36	42.2	4.1	58857	21	AA058471
37	42	4.1	4097	16	AA094255
38	41.8	4.1	990	20	AA084455
39	41.8	4.1	990	22	AA083338
40	41.8	4.1	2562	22	AA026469
41	41.8	4.1	8169	19	AA026609
42	41.8	4.1	44377	18	AA078508
43	41.8	4.1	44377	18	AA080414
44	41.6	4.1	492	22	AA057489
45	41.6	4.1	492	22	AA026979

ALIGNMENTS

RESULT	ID	AA04666	standard; DNA; 1026 BP.
XX	XX	AA04666;	
XX	XX	04-JUL-2001 (first entry)	
XX	XX	Thermus replication protein, Rept encoding DNA from pTSP45S plasmid.	
XX	XX	Replication protein; Rept; partition protein; Para; pTSP45S plasmid;	
XX	XX	kanamycin-resistance gene; thermophilic transformation; Ori;	
XX	XX	replication origin; ds.	
XX	XX	Thermus sp.	
XX	XX	Key	Location/Qualifiers
XX	XX	CDS	1..1026
XX	XX	/*tag= a	/*product= "Replication protein, Rept"
XX	XX	/codon= (Seq:GTC, aa:Met)	
XX	XX	US6207377-B1.	
XX	XX	27-MAR-2001.	
XX	XX	14-AUG-1998; 98US-0134246.	
XX	XX	14-AUG-1998; 98US-0134246.	
XX	XX	(NEW) NEW ENGLAND BIOLABS INC.	

groEL-1 gene codin
groEL-1 gene. Str
Pseudomonas aerugi
HSV-2 strain SB5 C
Zea mays DNA fragm
S. aggregatum PKS
GFP-HSFI fusion ge
Pseudomonas aerugi
Drosophila melanog
Pseudomonas aerugi
Synthetic human Fa
cDNA sequence of a
Bordetella pertuss
Sequence encoding
Bordetella pertuss
Streptomyces nours
Streptomyces nours
Gene conferring te
Streptomyces fireo
Streptomyces roseo
Nucleotide sequenc
DNA sequence used
Thiomonas cuprina
H. vulgare nicotia
Human Na+/H+ excha
DNA encoding novel
Nucleotide sequenc
Endothelial nitrog
AA094255
Human secreted pro
Ancestral HIV-1 gr
Actinmadura hibis
Platenolide syntha
Human foetal liver
Probe #5445 for ge

PS Example 1: Fig 3: 32pp; English.

XX The present sequence is *Thermus* plasmid pTsp45S DNA. The open reading
CC frame of pTsp45S plasmid is the replication protein Rept, partition
CC for thermophilic plasmid replication.
CC The invention relates to *Thermus* sp. replication protein Rept, partition
CC protein Para and their corresponding DNA molecules which relates to
CC recombinant DNA molecules encoding plasmid DNA replication origins in
CC *Thermus*, as well as to shuttle vectors which contain the same. The
CC invention also relates to method useful for cloning *Thermus* sp. plasmid
CC genes which comprises inserting plasmid DNA comprising a *Thermus* sp.
CC origin of replication (Ori) into a recombinant plasmid comprising a
CC thermostable kanamycin-resistance gene and an *Escherichia coli* Ori, to
CC produce a cloned recombinant plasmid. This cloned recombinant plasmid is
CC transformed with an *E. coli* host cell, and *E. coli* host cell cultured
CC for the expression of cloned recombinant plasmid. The cloned recombinant
CC plasmid isolated from *E. coli* host cell is then transformed with
CC *Thermus* sp. host cell and *Thermus* sp. host cell is cultured. Thus
CC *Thermus* sp. plasmid genes are cloned. These plasmid DNAs are used for
CC thermophilic transformation.

Sequence 5849 BP; 1377 A; 1501 C; 1670 G; 1301 T; 0 other:

Query Match 100.0%; Score 1026; DB 22; Length 5849;
Best Local Similarity 100.0%; Pred. No. 3.2e-230;
Matches 1026; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtgaagaacgaacaaacccctcttgaagaagcttacgaagctttagaagaaacccacgac 60
DB 4157 GTGAAGAAGCAAAAAACCTTTCTTGAAGAGCTTTAGAGAGCTTTAGAGAAACCCACGAC 4098
QY 61 aacaccgatgacacatgaggggtcagaatgaggggtcagaagactctcttggccaccgac 120
DB 4097 AACACCGATGCCATGAGGGGCTAGATAGGGGGCTCAGAGACTCTTCTTGGCCACGAC 4038
QY 121 ccccccagatgaggtgctccgaacatcgctcgcaagggcttatacaacaaagag 180
DB 4037 CCCCCCAGATGAGGTGCTCCGAATCCGCTCGCAAGGGCTTTCATACCAAAAAAGAG 3978
QY 181 gcaactagatgtcttaccacgaagaacacatgaggtcttctctctctgttgggccc 240
DB 3977 GCACCTAGATGCTTCTTACCCGAGAAAGACATGAGGCTTCTTCTTCTGTTGGGGCC 3918
QY 241 ccccccataccacacacgtacacccccctgttggaaatgataatgtaacgcctcagagcag 300
DB 3917 CCCCCCTATACACACACGTACACCCCGCTGGGAATGATGTAAGCCGTCAGAGAGGG 3858
QY 301 cctcagaagcttctggaactcctcagaagattgcccgcgtccacatccctcagcgaac 360
DB 3857 CCTCAGAAGCTTCTGGAACCTCTCCAGAGATTGCCCGCTCCACATCCCTTACGGCAAC 3798
QY 361 cggagagctctggaagaagtlyggaagcgtctcattcattgctccctcgaagatgtggcc 420
DB 3797 CGGAGACTCTGAGAAAGGAGGAGGAGGCTCTCTCATGCTCCCTGAGAGATTGGCC 3738
QY 421 cttaaccttgggttcaacccggcagacgcttcaacgctctggaagaagtctcttggaaaaag 480
DB 3737 CTTAACCTTGGGGTTCACCCGGCAGACCGCTCAGGCTGGAAGAGGCTTGAAGAAAAAG 3678
QY 481 ggcctgtgacacagcagctctcaccacaaacgctcaacaggagagcgccgacatcgac 540
DB 3677 GGCTGTGTGGCCACGACGCTCTTCAACAAACGCTCACGGGAGGCGCCGGCCATCGGC 3618
QY 541 acccttggcgctcggcgtgagcaggaagaagcagagctcaccctggaagctacatc 600
DB 3617 ACCCTTGGCGCTCGCGGCTGAGGCGCAGGGAAGCCAGGCTCACCTGAGCAGCATCATC 3558
QY 601 taccctggaggaacgctcgccctagacatgccaagcggtgtctctcctcaatggtgc 660
DB 3557 TACCCCTGGAGGAACCTCGCCTTAGACATGGCCAAACGCGCTGCTCTTCAACTGGGTC 3498
QY 661 aagcgctaccagacacgaatcgcgccacacctgtgactgtgtcctctggctcag 720

DB 3497 AAGGCTACAGACGACCGAATCCGCCACCCCTGAGAGTGTGCTCTCGGCTCAG 3438
QY 721 gggaaaggtatgtccacaacacacagacgctggccgttgaactggcctcactgtgc 780
DB 3437 GGGAAAGGGTGTGCTCCCAACCAAGACCGTGGCCGTGAGCTGGGCTCATCTGGGTC 3378
QY 781 ccccccgaagtgtgaggttccaaactccggcccttatccctcattgtctacgactt 840
DB 3377 CTCCCGAAGTGAAGGCTTCAAACTCCGGCCCTTATACCTCATCTGATGATATT 3318
QY 841 gccgactcctagatgacacgctgcttcaagagcttctatcagagctgtgtgtgtgtg 900
DB 3317 GCGGATCTCTATGATGACGCTGCTTCAAGAGCTTCTATGACAGCTTGTGTGGCTGTG 3258
QY 901 gccaggggtgaactcccgcgcaatatctattgtccgtccctaaatcaggtatcagat 960
DB 3257 GCCAGGGGTGAACCTCCCGCAATATCTATTGCTCTATGCGGGCTTATCCGAGAT 3198
QY 961 tacagagatggccatctgacacagcagagcgcttacctagtgaagaccccaagagggc 1020
DB 3197 TACAGGATGGCCATCTGTGACACGACCGGAGGCTTACTAGTGAAGACCTCAAGAGGCC 3138
QY 1021 tcctga 1026
DB 3137 TCCTGA 3132

RESULT 3
AAQ37036
ID AAQ37036 standard; DNA; 615 BP.
XX
AC AAQ37036;
XX
DT 05-JUL-1993 (first entry)
XX
DE SOD-T gene.
XX
KW Super oxide dismutase; heat resistant; medicine; inflammation;
KW X-ray; enzyme electrodes; bioreactor; ss.
XX
OS *Thermus aquaticus*.
XX
PN JP05030976-A.
XX
XX 09-FEB-1993.
PD
XX 26-JUL-1991; 91JP-0208747.
PF
XX 26-JUL-1991; 91JP-0208747.
PR
XX (YONS-) YOTSUBA NYUGYO KK.
PA
XX WPI: 1993-087958/11.
DR P-PSDB; AAR32376.
DR
XX

PT Gene, of and plasmid contg. heat resistant super:oxide dismutase
PT - for treating inflammation, X-ray protection, enzyme electrodes
PT and bio:reactor enzymes
PS
XX Claim 1: Page 7; 9pp; Japanese.
XX
XX The sequence shown is that isolated from *Thermus aquaticus* genomic
CC DNA and encodes a heat resistant super oxide dismutase enzyme. The
CC gene is useful for medicines for inflammation and as a protection
CC against X-rays, as a reagent for research in enzyme electrodes or as
CC an enzyme for a bioreactor.
XX

Sequence 615 BP; 125 A; 223 C; 183 G; 84 T; 0 other:

Query Match 5.3%; Score 54.4; DB 14; Length 615;
Best Local Similarity 47.5%; Pred. No. 0.0023;


```

FT      /note= "encoded protein shown in AAW72095, Xaa=
FT      unspecified amino acid"
FT      CDS      complement (5962..9306)
FT      /tag= b
FT      /product= "ORF#2 protein"
FT      /note= "encoded protein shown in AAW72096"
FT      9246..11195
FT      /tag= c
FT      /product= "ORF#3 protein"
FT      /note= "encoded protein shown in AAW72097"
FT      11657..15199
FT      /tag= d
FT      /product= "ORF#4 protein"
FT      /note= "encoded protein shown in AAW72098"
FT      15252..16265
FT      /tag= e
FT      /product= "ORF#5 protein"
FT      /note= "encoded protein shown in AAW72099"
FT      complement (16548..17621)
FT      /tag= f
FT      /product= "ORF#6 protein"
FT      /note= "encoded protein shown in AAW72100"
FT      18521..19936
FT      /tag= g
FT      /product= "ORF#7 protein"
FT      /note= "encoded protein shown in AAW72101"
FT      20015..21034
FT      /tag= h
FT      /transl_except= (pos: 21033..21034, aa: Gly-Ser)
FT      /product= "ORF#8 protein"
FT      /note= "encoded protein shown in AAW72102"
FT      W09820016-A1.
FT      14-MAY-1998.
FT      31-OCT-1997; 97WO-US20016.
FT      09-JUN-1997; 97US-0049018.
FT      04-NOV-1996; 96US-0030279.
FT      (SMK ) SMITHKLINE BEECHAM CORP.
FT      Chan JY, Dabrowski-Amara CE, Delvecchio AM, Dillon SB;
FT      Esser KM, Leary JJ;
FT      WPI: 1998-286847/25.
FT      P-PSDB: AAW72095, AAW72096, AAW72097, AAW72098, AAW72099, AAW72100,
FT      AAW72101, AAW72102.
FT      Herpes simplex virus type-2 sequences - useful in, e.g., prevention
FT      and treatment of infection or inducing immunological response in
FT      mammal
FT      Claim 1; Page 350-359; 74BPP; English.
FT      This sequence represents a Herpes simplex virus type-2 (HSV-2) DNA
FT      sequence of the invention. This sequence was isolated from HSV-2 strain
FT      SB5 (deposited as ATCC VR-2546), is designated Contig ID 10, and encodes
FT      8 HSV-2 proteins. The proteins can be used for the treatment or
FT      prevention of disease, to induce an immunological response in a mammal or
FT      to identify inhibitors, activators or novel antivirals. Antagonists of
FT      the proteins can be used to inhibit a viral polypeptide. The DNA sequence
FT      or a vector containing it can also be used to induce an immunological
FT      response in a mammal.
FT      Sequence 21034 BP; 3154 A; 7419 C; 7234 G; 3227 T; 0 other;

```

Query Match 4.7%; Score 48.6; DB 19; Length 21034;
 Best Local Similarity 43.9%; Pred. No. 0.12;
 Matches 207; Conservative 0; Mismatches 264; Indels 0; Gaps 0;

```

QY      355  ggaacccggagctctggaagaagtgggagcgtgtcttattatgttccctggagatg 414
QY      2012  gccatgacagctctacagaccttggtgggtggccgacgtgattctcagggccgc 1953
QY      415  ttggccctcaacctgggggtcaaccgcgcagacccgtccacgcctggaagaagttccttgg 474
QY      1952  cgggcttgccgcccacgcccaggggacacatgcccctctatgatccgccgcccgcgcg 1893
QY      475  aaaaaggccctgtggtccacgcagcttccttcaaccaaccgtaacgaggagcgcgggcc 534
QY      1892  acgaccgctgcttgcggaaagacctcgggggacagagctggggtcgggggcctgtgcac 1833
QY      535  atcgagaccttggcgcgttcgcgtgaggaagcgaaggaagcgaagccttgagcgcgc 594
QY      1832  gccgaaggacggccgggacatttccccagcggcgccctgagagactgggccaag 1773
QY      595  tacatctacccttgaggaacctcgccctagacatgtgccaagcgcgtgtcttccttaac 654
QY      1772  gmcattggcgccacgacgacgcccgcgcagacacttgagggccgctgcgcacgtgacg 1713
QY      655  tgggtcaaggcctaccaggaacgaatcgcccccacccttgagcgtgtgtctcttgg 714
QY      1712  gccgaagatggcgccacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1653
QY      715  gctcaagggaagaagggtgtagtgcaccacacgaagcgtgacgtgtgacctggcctcacc 774
QY      1652  gcggcgctggaccgctccagaaacccgcggaatttcagctgagctgacgtgcgcgcctg 1593
QY      775  ctgtctctcccgaggtgagcgttccaaactccgcgccttatacccttc 825
QY      1592  caggcctggcgccgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 1542

```

RESULT 6
 AAV62176/c
 ID AAV62176 standard; DNA; 117213 BP.
 AC AAV62176;
 XX 13-JAN-1999 (first entry)
 DT
 DE HSV-2 strain SB5 Contig ID 15 DNA sequence.
 XX
 XX HSV-2 strain SB5; immunological response induction; therapy;
 KW antiviral identification; viral protein inhibitor; ss.
 XX
 OS Herpes simplex virus type 2.
 XX
 XX
 FH Key Location/Qualifiers
 FT 755..1297
 FT CDS
 FT /tag= a
 FT /product= "ORF#1 protein"
 FT /note= "encoded protein shown in AAW72170"
 FT 1170..2174
 FT /tag= b
 FT /product= "ORF#2 protein"
 FT /note= "encoded protein shown in AAW72171"
 FT 2229..2930
 FT /tag= c
 FT /product= "ORF#3 protein"
 FT /note= "encoded protein shown in AAW72172"
 FT complement (3130..3735)
 FT /tag= d
 FT /product= "ORF#4 protein"
 FT /note= "encoded protein shown in AAW72173"
 FT complement (3802..6447)
 FT /tag= e
 FT /product= "ORF#5 protein"
 FT /note= "encoded protein shown in AAW72174"
 FT 6017..8482
 FT /tag= f
 FT /product= "ORF#6g protein"
 FT /note= "encoded protein shown in AAW72250"

```
FT CDS 6026..8482
FT /*tag= g
FT /product= "ORF#6f protein"
FT /note= "encoded protein shown in AAW72249"
FT 6055..8482
FT /*tag= h
FT /product= "ORF#6e protein"
FT /note= "encoded protein shown in AAW72248"
FT 6167..8482
FT /*tag= i
FT /product= "ORF#6d protein"
FT /note= "encoded protein shown in AAW72247"
FT 6296..8482
FT /*tag= j
FT /product= "ORF#6c protein"
FT /note= "encoded protein shown in AAW72246"
FT 6326..8482
FT /*tag= k
FT /product= "ORF#6b protein"
FT /note= "encoded protein shown in AAW72245"
FT 6446..8482
FT /*tag= l
FT /product= "ORF#6a protein"
FT /transl_except= (pos: 7400..7402, aa: Ala-Ala)
FT /transl_except= (pos: 7481..7486, aa: Ile)
FT /note= "encoded protein shown in AAW72244"
FT 8457..9347
FT /*tag= m
FT /product= "ORF#7 protein"
FT /note= "encoded protein shown in AAW72175"
FT complement (9604..11855)
FT /*tag= n
FT /transl_except= (pos: 11635..11636, aa: Ala)
FT /product= "ORF#8 protein"
FT /note= "encoded protein shown in AAW72176"
FT complement (11905..14508)
FT /*tag= o
FT /product= "ORF#9b protein"
FT /note= "encoded protein shown in AAW72222"
FT complement (11905..14520)
FT /*tag= p
FT /product= "ORF#9a protein"
FT /note= "encoded protein shown in AAW72223"
FT 14399..15802
FT /*tag= q
FT /product= "ORF#10 protein"
FT /note= "encoded protein shown in AAW72177"
FT complement (15996..16286)
FT /*tag= r
FT /product= "ORF#11 protein"
FT /note= "encoded protein shown in AAW72178"
FT complement (16202..18064)
FT /*tag= s
FT /product= "ORF#12 protein"
FT /note= "encoded protein shown in AAW72179"
FT complement (18105..19661)
FT /*tag= t
FT /product= "ORF#13 protein"
FT /note= "encoded protein shown in AAW72180"
FT complement (19415..20074)
FT /*tag= u
FT /product= "ORF#14 protein"
FT /note= "encoded protein shown in AAW72181"
FT 20155..21453
FT /*tag= v
FT /product= "ORF#15 protein"
FT /note= "encoded protein shown in AAW72182"
FT complement (21326..22291)
FT /*tag= w
FT /product= "ORF#16 protein"
FT /note= "encoded protein shown in AAW72183"
FT complement (22546..24654)
FT /*tag= x
FT CDS

FT CDS 6026..8482
FT /*tag= g
FT /product= "ORF#17 protein"
FT /note= "encoded protein shown in AAW72184"
FT 24684..25955
FT /*tag= y
FT /product= "ORF#18 protein"
FT /note= "encoded protein shown in AAW72185"
FT complement (26295..27251)
FT /*tag= z
FT /product= "ORF#19 protein"
FT /note= "encoded protein shown in AAW72186"
FT complement (27630..31754)
FT /*tag= aa
FT /product= "ORF#20b protein"
FT /note= "encoded protein shown in AAW72225"
FT complement (27630..31784)
FT /*tag= ab
FT /product= "ORF#20a protein"
FT /note= "encoded protein shown in AAW72224"
FT complement (32067..32735)
FT /*tag= ac
FT /product= "ORF#21 protein"
FT /note= "encoded protein shown in AAW72187"
FT 33140..34984
FT /*tag= ad
FT /product= "ORF#22a protein"
FT /note= "encoded protein shown in AAW72226"
FT 33386..34984
FT /*tag= ae
FT /product= "ORF#22b protein"
FT /note= "encoded protein shown in AAW72227"
FT complement (35205..37721)
FT /*tag= af
FT /product= "ORF#23 protein"
FT /note= "encoded protein shown in AAW72188"
FT complement (38058..39188)
FT /*tag= ag
FT /product= "ORF#24 protein"
FT /note= "encoded protein shown in AAW72189"
FT 39090..39935
FT /*tag= ah
FT /product= "ORF#25 protein"
FT /note= "encoded protein shown in AAW72190"
FT 40216..41973
FT /*tag= ai
FT /product= "ORF#26 protein"
FT /note= "encoded protein shown in AAW72191"
FT 42206..44178
FT /*tag= aj
FT /transl_except= (pos: 44063..44064, aa: Lys)
FT /product= "ORF#27 protein"
FT /note= "encoded protein shown in AAW72192"
FT complement (44853..47297)
FT /*tag= ak
FT /product= "ORF#28 protein"
FT /note= "encoded protein shown in AAW72193"
FT 47122..47338
FT /*tag= al
FT /product= "ORF#29 protein"
FT /note= "encoded protein shown in AAW72194"
FT complement (47305..49662)
FT /*tag= am
FT /product= "ORF#30 protein"
FT /note= "encoded protein shown in AAW72195"
FT complement (50035..51666)
FT /*tag= an
FT /product= "ORF#31 protein"
FT /note= "encoded protein shown in AAW72196"
FT complement (51701..53575)
FT /*tag= ao
FT /product= "ORF#32 protein"
FT /note= "encoded protein shown in AAW72197"
FT 54393..58115
FT /*tag= ap
FT CDS
```

FT	/product= "ORF#33 protein"
FT	/note= "encoded protein shown in AAW72198"
FT	complement (58060..58977)
FT	/*tag= ag
FT	/product= "ORF#34 protein"
FT	/note= "encoded protein shown in AAW72199"
FT	complement (58970..60760)
FT	/*tag= ar
FT	/product= "ORF#35 protein"
FT	/note= "encoded protein shown in AAW72200"
FT	60759..61151
FT	/*tag= as
FT	/product= "ORF#36 protein"
FT	/note= "encoded protein shown in AAW72201"
FT	61241..62071
FT	/*tag= at

Query Match	4.78;	Score 48.6;	DB 19;	Length 117213;
Best Local Similarity	43.98;	Pred. No. 0.19;		
Matches 207; Conservative	0;	Mismatches 264;	Indels 0;	Gaps 0;

Db	68405	GGCATGAGAGACTCAGACCCCTGGGGTGGGGGCGCCGACCTGGTAGTCCAGGGCGCC	68344
Oy	415	ttggacctaaacttgggtgtcaaccggcagaaccgtccacgctgtgaagaagtctctttag	474
Db	68345	CGGGCTTGCGCCACGGCGGAGGGGAGACATGCGCCCTCATGATGCGCGGGCCCGGG	68286
Oy	475	aaaagggccttgttgcacacgacgtctctaaccaaccgttaaggggagcgcggcc	534
Db	68285	ACGACCGCTGCTCGGGAAAGCCTTCGGGGGCGACAGAGCTGGGTGGGGGCTGCTGCAC	68226
Oy	535	atcggcaccccttggcgccgttcgctgtgaagccaaaggaaagcagcgtcaacctgtgacgac	594
Db	68225	GCCGGAAGGACGCGCGGGGACCATTCGCCACGCGGGCGCGCCTCAGAGAGCTGGGCAAG	68166
Oy	595	tacatctaccccttgaagaaactctgcctcaagcatalygcacaaagcgtgtctctcctaac	654
Db	68165	GTCACTCGCGCGCACCGCGACGCCGCCCGACGCAACTTGAGGCCGCGCTGCCCACTGACG	68106
Oy	655	ttggtcaagagcctacaaagaccagaaatctcgcccccaaccttgcagctgtgttctcttgg	714
Db	68105	GCGAAGATGCGGGGCCACGCGCCGCCCGGGGCAACAGACGATGGGGCGCGCCGCTCGAG	68046
Oy	715	gtctcaggggaaaaaggtgtatgtcccaacaacaagaccgttgccttlytaacttggcctatc	774
Db	68045	GCGGCGTGGACCGCGTGGAAACCGCGCCGACGTTCGACGTGGTGCAGCTGCCCGCGCTG	67986
Oy	775	ctgtgtctccccgaagtgtgaagcttccaaactcccgcccttatcacctc	825
Db	67985	CAGGGCTGGGGGCGACACGCTACACACCCTCCGGGACCTTCGGAAACGCG	67935

Accession ID	Sequence
RESULT 7	
AAD25519/c	
XX	
XX	AAD25519 standard; DNA: 154746 BP.
AC	
XX	AAD25519;
DT	
XX	26-MAR-2002 (first entry)
XX	
De	
XX	Human herpesvirus 2 complete DNA genome.
KW	
KW	Human herpesvirus 2; cytostatic; cancer; immunosuppressive; virucide;
KW	antibacterial; fungicide; protozoacide; antirheumatic; antiinflammatory;
KW	antibacterial; rheumatoid arthritis; neuroprotective; multiple sclerosis
KW	immune response; vasotropic; vaccine; gene therapy; autoimmune disease;
KW	vasculitis; ds.
XX	
XX	
OS	Human herpesvirus 2.
XX	
XX	
NN	MO200176643-A1.

XX 18-OCT-2001.
PD
XX
PE 06-APR-2001; 2001WO-US11372.
XX
PR 07-APR-2000; 2000US-195680P.
XX
PA (BAYU) BAYLOR COLLEGE MEDICINE.
XX
XX
PI Orson FM, Kinsey BM, Bhogal BS;
XX
DR WPI; 2002-066308/09.
XX
XX
PT Composition for oral delivery of vaccines, comprises expression vector
PT containing antigenic genomic sequence, bound to aggregated
PT protein-polycationic polymer conjugate or suspension
PS Disclosure: Page 90-132; 145pp; English.
XX

The invention relates to a composition comprising an expression vector bound to an aggregated protein-polyanionic polymer conjugate or suspension. The expression vector contains a promoter polynucleotide sequence operatively linked to a polynucleotide sequence encoding an antigen which is a fragment of a gene or genome associated with an infectious disease, cancer and autoimmune disease such as rheumatoid arthritis, vasculitis, and multiple sclerosis, pathogenic genomes consisting of bacterium, fungus, protozoa and virus such as human immunodeficiency virus (HIV), herpes simplex virus (HSV), hepatitis C virus (HCV), influenza and respiratory syncytial virus (RSV), and optionally comprising a nucleotide sequence encoding a cytokine (or a cytokine expression vector), is useful for inducing an immune response (systemic and/or mucosal) in an organism. The cytokine expression vector contains a sequence for granulocyte macrophage-colony stimulating factor (GM-CSF) or interleukin-12 (IL-12). The polynucleotide sequences encoding the antigen and the cytokine are under transcriptional control of same or different promoter polynucleotide sequences. The expression vector, as a DNA vaccine is useful for treating a condition in an organism. The present sequence is human herpesvirus 2 complete DNA genome related to the invention.

Sequence 154746 BP; 23003 A; 54218 C; 54701 G; 22824 T; 0 other;

Query Match	4.78;	Score 48.6;	DB 24;	Length 154746;
Best Local Similarly	43.98;	Pred. No. 0.2;		
Matches 207; Conservative	0;	Mismatches 264;	Indels 0;	Gaps 0.

QY	355	ggcaacgggagctctcgggaggaagltgggagcgtgcgtcttcaatgctccctcgagatg	414
Db	77292	GCATCGACAGCAGTCAACGACCTGGGGTCTGGGGCGCCGACCTGGTATGTCAGGGCGCC	77233
QY	415	ttggccctcaaccttgggggtcaaccggtacagctlcaacgcttggaaagagctcttgg	474
Db	77232	CGGGCGCTGGCCACGCGGAGGGGAGCCATGCGCGCCCTCATCATGATGCCGGCGCCCGC	77173
QY	475	aaaagggcctcgttggccaacgagctctcttcaacaacgctcaacgvggagcgccg	534
Db	77172	ACGACCGCTGCTCGGGAAGCTCTCGCGGGACACAGGCTGGTTCCGGGGCTGCTGCAC	77113
QY	535	atcggcaaccttttggccgtccggtcgtggtgacgaagaaacagagctcaacctggagcag	594
Db	77112	GCGGAAGGAGCGCGCGGAGACATTCCCCACGCGCGGCCCTCCAGGAGCTGGGCAAG	77053
QY	595	tacatctaacctctgtagaagaaactcgtccctctagatctgccaagcggtctctctcaac	654
Db	77052	GTCATCGCGCGCACCGACGCGCGCGCGACAACTTGAAGCCGCCCTCCACACCTGACG	76993
QY	655	tgggtcaagcgctacagagacacaggaatccggcccaaccttgaacgtgtgtctctctgg	714
Db	76992	GGGAAGATGGCGGCCACGCGCGCCGGGGACGACGACGCGCTGGGCCGCCGCGCTCAG	76933
QY	715	gtctcaggggaaaagtgatctgccacaacaagaccttgcgtttgacctgggctctatc	774

Db 76932 GCGGCGGTGACCGCGCGAGACCGCGCGAGTTCGAGCTGCGAGCTCGCCGCTG 76973
 OY 775 ctgtgtctcccccgaagtgtgagcgttccaaatcccgccctatcaccttc 825
 Db 76972 CAGGCGCTGGCGGACGACGCGCTACACCCCGGAGACTTCGCGAAGCCG 769822

RESULT 8

AAAT1428
 ID AAA71428 standard; DNA: 6094 BP.

XX AAA71428;

AC 01-DEC-2000 (first entry)

DE Expression vector pSEX1114 encoding DNA.

XX Expression vector; antibody binding protein; monoclonal antibody; Neo-R;
 KW B lymphocyte; myeloma cell; hybridoma; protein G; b1a protein; ds.

XX Synthetic.

XX Location/Qualifiers

FT Key 682..1785

FT /tag- a

FT /product- "G protein"

FT complement (3402..4190)

FT /tag- b

FT /product- "Neo-R"

FT complement (4645..5505)

FT /tag- C

FT /product- "b1a"

DE19900635-A1.

PD 13-JUL-2000.

XX 11-JAN-1999; 99DE-1000635.

XX 11-JAN-1999; 99DE-1000635.

XX 11-JAN-1999; 99DE-1000635.

PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.

XX Breittling F, Poustka A, Moldenhauer G;

PI Breittling F, Poustka A, Moldenhauer G;

XX WPI: 2000-499832/45.

DR P-PSDB: AAB10432, AAB10437, AAB10438.

DR Selecting monoclonal antibodies, by expressing them on the surface of
 hybridomas attached to antibody-binding protein, then reaction with

antibody library

antibody library

Claim 8; Fig 1; 22pp; German.

This invention describes a novel method for the selection of monoclonal
 antibodies (Mab) which comprises (i) fusing B lymphocytes with myeloma
 cells to produce antibody-producing hybridomas such that the antibodies
 are presented at the surface of the hybridomas by an antibody-binding
 protein (I); and (ii) binding the antibody to antigens (Ag). The
 invention also describes antibody-binding proteins (I) that comprise a
 combination of the signal peptide of a murine immunoglobulin (Ig) kappa
 chain or a murine MHC (major histocompatibility complex) Class I k(k)
 molecule; an antibody-binding site of proteins A, G, I or Ig, and the
 transmembrane domain of PDGR (platelet-derived growth factor receptor)
 or CD52. The method is used to select Mab with specificity for particular
 antigens. Mab can be selected without separate culture of hybridomas,
 on the basis of strength of affinity for a particular antigen. Complex
 mixtures of hybridomas can be used for selection, reducing the time and
 cost involved in Mab selection. This sequence represents the expression
 vector pSEX1114 encoding protein G, Neo-R and the b1a protein which is
 used in the method of the invention.

SO Sequence 6094 BP: 1465 A; 1684 C; 1599 G; 1346 T; 0 other;

Query Match 4.68; Score 47; DB 21; Length 6094;

Best Local Similarity 44.1%; Pred. No. 0.22; Indels 0; Gaps 0;

Matches 197; Conservative 0; Mismatches 250;

OY 355 ggcacacgggaagctctgtgaggaagtggtggagcgtctgtccatgtcccttggaagtg 414
 Db 655 gggagagcccaagctgtgacgtgtgacgtgtgacccctgtgctctgtctgtgag 714
 OY 415 ttggccctcaaccctgtgggtgtaccccggaagacgtgtccacgtggaagaagtccttgag 474
 Db 715 gccgccttgcccgacatcagaccgcggtgggtggcccaagaagagaccgcccgagag 774
 OY 475 aaaaagggcctgtgtgcccacgacgtctctcccaaacggtcaacgggagcgcgggccc 534
 Db 775 cccaagagagaggtgtgacccatcaagggcaccgtatcctacgacgggcaagaccagacc 834
 OY 535 atcgagacccttggtggcgtctgcgtgaagccagggaaagcaggtcacccttgagacgac 594
 Db 835 gccgaagttcaaggagcacccttgagagggccacgcggtgaggtcctacggtacgcgagcgc 894
 OY 595 tacatctacccctgtgaggaacctgcctcagacatgtgccaaggggtgtctctcctcaac 654
 Db 895 ctgaagaagagacaacgagcaggtacacgtgtgagcgtgtgacgaaggggtaccacccctgaac 954
 OY 655 ttgggtcaagcctcaccagagacacacggtatccgcccaccttgagcgtgtgtctctgtg 714
 Db 955 atcaagttcgcgcggaaggaaggaaccccgaggaagcccaaggaaggtgtgacatcaag 1014
 OY 715 gctcaggggaaaggggtgtatgtcccaacacgaacacgtgtgctgtgacttggtggccctc 774
 Db 1015 gccacactgtatctacgcgcgacgaagaccacgacccgagttcaaggacacttcgag 1074
 OY 775 ctgtctctcccgaggtgtgagcgcttc 801
 Db 1075 gaggcacacgcgaggtgacgtaccgtac 1101

RESULT 9

AAV34317/C

ID AAV34317 standard; DNA: 1473 BP.

XX AAV34317;

AC 29-JAN-1999 (first entry)

DE Human secreted protein gene 13 clone HNFcv70.

XX Human; secreted protein; fusion protein; gene therapy; protein therapy;

KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;

KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;

KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;

KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;

KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;

KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;

KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.

OS Homo sapiens.

PN WO9840483-A2.

PD 17-SEP-1998.

XX 12-MAR-1998; 98WO-US04858.

XX 19-DEC-1997; 97US-0068368.

PR 14-MAR-1997; 97US-0040710.

PR 14-MAR-1997; 97US-0040762.

PR 30-MAY-1997; 97US-0048100.

PR 30-MAY-1997; 97US-0048189.

PR 30-MAY-1997; 97US-0048357.

PR 30-MAY-1997; 97US-0050934.
 PR 06-JUN-1997; 97US-0048970.
 PR 05-SEP-1997; 97US-005765.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Ferrie AM, Fischer CL, Gentz RL, Greene JM, Kyaw H;
 PI Li H, Li Y, Moore PA, Rosen CA, Ruben SM, Soppet DR;
 PI Wei YF, Young PE, Zeng Z;
 XX
 DR WPI: 1998-520811/44.
 DR P-PSDB: AAW75227.
 XX
 PT Isolated human poly:nucleotide(s) encoding secretory peptide(s) -
 PT used to develop products for the diagnosis and treatment of e.g.
 PT inflammation, cancers, CNS disorders or immune system disorders
 XX
 PS Claim 1; Page 144-145; 201pp; English.

CC This sequence represents a nucleic acid molecule which encodes a
 CC secreted human protein. The gene number, and the clone it is derived
 CC from, are detailed in the descriptor line. The gene can be used to
 CC generate fusion proteins by linking to the gene to a human immunoglobulin
 CC Fc portion (e.g. AAV34277) for increasing the stability of the fused
 CC protein as compared to the human protein only.
 CC The invention relates to 28 novel genes and their fragments (nucleic
 CC acid sequences: AAV34286-V34325; amino acid sequences AAW75196-W75235)
 CC which are useful for preventing, treating or ameliorating medical
 CC conditions e.g. by protein or gene therapy. Also, pathological
 CC conditions can be diagnosed by determining the amount of the new
 CC polypeptides in a sample or by determining the presence of mutations in
 CC the new polynucleotides. Specific uses are described for each of the 28
 CC polynucleotides, based on which tissues they are most highly expressed in
 CC (see AAV34286 for described uses).
 CC
 XX

SQ Sequence 1473 BP; 257 A; 476 C; 491 G; 249 T; 0 other;

Query Match 4.5%; Score 46.2; DB 19; Length 1473;
 Best Local Similarity 42.5%; Pred. No. 0.23;
 Matches 299; Conservative 0; Mismatches 403; Indels 1; Gaps 1;

QY 314 tgaagctctccagagattgcccgtccacatccctcctgaagcagcgagcttga 373
 DB 823 TGGAACTCTCTGACTATTCAAATCTCAGGCCCAACAGCAGCAAGTTGAGT 764
 QY 374 ggaaggtgggagcgtgctcctcattgctcccttgagagatgttgcccaacccgag 433
 DB 763 TTCTGATATAGCTTGTCCAGCTGATTCACCACTCCAGCAACCCGATGACGAGGCTGCT 704
 QY 434 taaccgagagacgtccagcctggaagaagctccttgagaagaaggcctgtgtgcca 493
 DB 703 TCGCTCCCGGATCTGGAACCTGCTGATGTCCACGACCAACATACAGCTCCAGTAGCGG 644
 QY 494 ccgagctccttaaccaaacgctcaagggagggcgccgagccttgagc-acctttgggccc 552
 DB 643 GTCCTCCGGGATGGAGAAAGTCCCGGGCCAGGCTTAAGAGGCTGCTCCGGGCT 584
 QY 553 gtccggtctgagcagggaggaagcagctcacccttgagactatactaccctcggag 612
 DB 583 CCCAGAGAGCTCCAGAGGCTGTGTCGTCGAGCCGCGCAGGCTCCGGCGCTGCGACGAG 524
 QY 613 aaccctgccttagaatgagcagggcgtgctcctccttaactgggtgaagcctaccag 672
 DB 523 TCGTCAAGCTGTATACAGGCGTCCGCTCCGCTCCGCACTTCATCCAGGGGCTCGATC 464
 QY 673 gaccacggaatccgccacccttgagcgtgtgtcctctgagcctgaggaagaggtg 732
 DB 463 AGGTCCAGGTCTGACCCACCTGGAAGAACCCCTGAGCGCGGACAGAGTGTGAGCGT 404
 QY 733 atgcccacaacaaagacgttgccgttgacttgagcctcactcctgctcccgaggtg 792
 DB 403 GCGGCTGAGTCCGTTACCTCTTAAGTGGCCCTGTGAAGAAGAGTGTGCTCTCCGCGAGG 344

QY 793 gaggcttccaaactcccgccctatataccctcattgtctacgtacattgccgataccta 852
 DB 343 CTGCTCCGTCACCTCGGAGCGCATGTCGAGCGCGATAGTCTCTGTGTAGCGGACCCAG 284
 QY 853 gatgaccgtgcttcaagaagcttctatgcaaggtctgtggtggtggtggtggtgaa 912
 DB 283 CAGGTCCCTGTTCTTCGCGAGGTGAGGTGAAGTTGTCCCTGTGCGCCCAAGAGAGTA 224
 QY 913 ctcccgcgcaatactatttgccgtccttaattgaggttaccgagattacaagatgagc 972
 DB 223 GCTTACCTCTCTGCGGTGCGAGGCCCAAGTGGAGGCGAGAGCTTGGCGGAGCTGGGCGCC 164
 QY 973 catctgacacgacggagagcgtactagtgaaagaccctcaag 1015
 DB 163 TGGCAGACGCCACGCGCAACACAGACTCATCTGCTCATGAGG 121

RESULT 10

AAQ22482
 ID AAQ22482 standard; DNA; 1620 BP.

AAQ22482;

29-JUL-1992 (first entry)

DE groEL-1 gene coding region.

XX Heat shock protein; HSP18; ss.

XX Streptomyces albus.

XX Key Location/Qualifiers

FT CDS 1..510 /note="groEL-1 gene encoding HSP18"

PN W09204452-A.

PD 19-MAR-1992.

PF 03-SEP-1991; 91WO-FR00701.

PR 10-SEP-1990; 90FR-0011186.

PA (INSP) INST PASTEUR.

PI Mazodier P, Guglielmi G;

DR WPI: 1992-114358/14.

DR P-PSDB: AAR22363.

PT Recombinant DNA contg. heat inducible promoter and heterologous

PT gene - also vectors, transformed cells and new heat shock

PT proteins of Streptococcus albus

PS Claim 18; Fig 8; 50pp; French.

CC The sequence is that of the coding region of the groEL-1 gene which

CC codes for the 18 kDa heat shock protein HSP18. The gene actually

CC modification to give the 18 kDa HSP18 protein.

CC See also AAQ22477-Q22486.

SQ Sequence 1620 BP; 287 A; 590 C; 539 G; 204 T; 0 other;

Query Match 4.4%; Score 45.6; DB 13; Length 1620;
 Best Local Similarity 43.1%; Pred. No. 0.33;
 Matches 219; Conservative 0; Mismatches 289; Indels 0; Gaps 0;

QY 289 gtcaagagcggcctcagaagctcttgaaactcctcagaagatgccgctcacatc 348

Db 1081 gaccgagagaagctccgagagcgcctcgcacagctggccgagcgtctgctgataccgc 1140
 QY 349 cccctacgcaacccgagcctctgagagaaggctgggagcgtctctatcatgtccctc 408
 Db 1141 gtcgagcgagcaccgagagctgagctgagagagcgcacacgctctcggagagccatc 1200
 QY 409 gagatgttgccctcaacctgggggtcaaccgagacgcgtccacgctggaaagatc 468
 Db 1201 tccgagaccgcgcgcgcgtctgagagagacatcgtctccggtgggtccgcgcgtg 1260
 QY 469 cttagagaaaggcgtctgctggcaccgacgtctctcaacaacgtcaacggggagcgc 528
 Db 1261 cagcgctcaagagctcctggagacacactcgtccgcacgcgcgagagccacccggtc 1320
 QY 529 cgggacatcgacaccccttggcgcgtccgcgtgagcagagaaagccacgctcacc 588
 Db 1321 ggcgtcgtccgcgcgcgcgtcgcagcgcgtgctgatacgcgagaaagccgcgc 1380
 QY 589 gacgactatcatcacccttgagaggaactcgcctcagacagggccaaagcgtgctcc 648
 Db 1381 gagggctcgtcatcaccacaaaggtgagcgtcgcagacaaaggccaaaggcttcaagcg 1440
 QY 649 tccaactgggtcaagagcctacacagacagacgaatccgcccacccctggagctgctc 708
 Db 1441 gccacgcgagagtaagcgacactgtgtaagcgcgctcatcgaccgcgtcaaggtcac 1500
 QY 709 cctcgggctcagggagaaagggtgatgcccacacaaagacggtgctgactggc 768
 Db 1501 cgcctcgcgcctggagaaagcgacactccatcgcctccctgctctcgtgagacgagac 1560
 QY 769 ctcatcctgtctcccccaggtgagc 796
 Db 1561 gtcgtcagagaagcgcgcgcgagagagagc 1588

RESULT 11

AAQ22485

ID AAQ22485 standard: DNA; 2668 BP.

XX AC AAQ22485;

XX DT 29-JUL-1992 (first entry)

XX DE groEL-1 gene.

XX KM Heat shock protein; HSP18; ss.

XX ST Streptomyces albus.

XX KEY Location/Qualifiers

XX CDS 849..2468 /tag= a /note= "Gro EL1 protein"

XX FT WO9204452-A.

XX PD 19-MAR-1992.

XX PE 03-SEP-1991; 91WO-FR00701.

XX PR 10-SEP-1990; 90FR-0011186.

XX PA (INSP) INST PASTEUR.

XX PI Mazodier P, Gugglielmi G;

XX DR WPI; 1992-114358/14.

XX DR P-PSDB; AAR22363.

XX PT Recombinant DNA contg. heat inducible promoter and heterologous

XX PT gene - also vectors, transformed cells and new heat shock

XX PT proteins of Streptococcus albus

PS Claim 18; Fig 8; 50pp; French.
 XX The sequence is that of the complete groEL-1 gene which codes for
 CC the 18 kDa heat shock protein HSP18. The gene actually codes for a
 CC 56 kDa protein but this is subjected to post translation modification
 CC to give the 18 kDa HSP18 protein. See also AAQ22477-Q22486.
 XX
 SQ Sequence 2668 BP; 443 A; 948 C; 929 G; 348 T; 0 other;

Query Match 4.48; Score 45.6; DB 13; Length 2668;
 Best Local Similarity 43.18; Pred. No. 0.37; Mismatches 289; Indels 0; Gaps 0;
 Matches 219; Conservative 0;

QY 289 gtcagagagcggtcctgaagctcttgaaactcctccacagagatggccctccacatc 348
 Db 1929 gaccgagagaagctcccgagagcgtctcgaagctggccggtcgtctgctgataccgc 1988
 QY 349 cccctacgcaacccgagagctctgagagaaggctgggagcgtctctatcatgtccctc 408
 Db 1989 gtcgagcgagcaccgagagctgagctgagagagcgcacacgctctcggagagccatc 2048
 QY 409 gagatgttgccctcaacctgggggtcaaccgagacgcgtccacgctggaaagatc 468
 Db 2049 tccgagaccgcgcgcgcgtctgagagagacatcgtctccggtgggtccgcgcgtg 2108
 QY 469 cttagagaaaggcgtctgctggcaccgacgtctctcaacaacgtcaacggggagcgc 528
 Db 2109 cagcgctcaagaggtcttgagacacactcgtccgcacgcgcgagagccacccggtc 2168
 QY 529 cgggacatcgacaccccttggcgcgtccgcgtgagcagagaaagccacgctcacc 588
 Db 2169 ggcgtcgtccgcgcgcgcgtcgcagcgcgtgctgatacgcgcgagaaagccgcgc 2228
 QY 589 gacgactatcatcacccttgagaggaactcgcctcagacagggccaaagcgtgctcc 648
 Db 2229 gagggctcgtcatcaccacaaaggtgagcgtcgcagacaaaggccaaaggcttcaagcg 2288
 QY 649 tccaactgggtcaagagcctacacagacagacgaatccgcccacccctggagctgctc 708
 Db 2289 gccacgcgagagtaagcgacactgtgtaagcgcgctcatcgaccgcgtcaaggtcac 2348
 QY 709 cctcgggctcagggagaaagggtgatgcccacacaaagacggtgctgactggc 768
 Db 2349 cgcctcgcgcctggagaaagcgacactccatcgcctccctgctctcgtgagacgagac 2408
 QY 769 ctcatcctgtctcccccaggtgagc 796
 Db 2409 gtcgtcagagaagcgcgcgcgagagagagc 2436

RESULT 12

AAS54366

ID AAS54366 standard: DNA; 1686 BP.

XX AC AAS54366;

XX DT 13-FEB-2002 (first entry)

XX DE Pseudomonas aeruginosa DNA for cellular proliferation protein #497.

XX KW Antisense; ds; prokaryotic cellular proliferation gene;

XX KW antibiotic; antibacterial; drug design.

XX OS Pseudomonas aeruginosa.

XX PN WO200170955-A2.

XX PD 27-SEP-2001.

XX PF 21-MAR-2001; 2001WO-US09180.

XX PR 21-MAR-2000; 2000US-191078P.

PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139753.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140981.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142390.
PR 09-JUL-1999; 99US-0142803.
PR 12-JUL-1999; 99US-0142920.
PR 13-JUL-1999; 99US-0142977.
PR 14-JUL-1999; 99US-0143542.
PR 15-JUL-1999; 99US-0143624.
PR 16-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145919.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.

PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148365.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0148368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151348.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0155458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159325.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 4.4%; Score 45.2; DB 21; Length 1290;
Best Local Similarity 47.2%; Pred. No. 0.38;
Matches 137; Conservative 0; Mismatches 153; Indels 0; Gaps 0;
Qy 479 agggcctgttgccaccgcagcgttcaccacaacgctcaacgggagcgccggccatcg 538
Db 76 aagtcacagcgcgccacgttgatgataacgcgccaagatcgcgcgcgccgctccg 135

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 29, 2002, 08:12:58 ; Search time 1687.79 Seconds

(without alignments)
8204.739 Million cell updates/sec

Title: US-09-664-186-4

Perfect score: 1026

Sequence: 1 gtgaagaacgaacaaacctt.....ccctcaagagagcctcctga 1026

Scoring table: IDENTITY NUC Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estlin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_estl:*
9: gb_estl:*
10: gb_est2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	54.6	5.3	925	12	CNS0091P
2	51.8	5.0	925	12	CNS0091P
3	50.6	4.9	637	10	BI960118 HYSME002
4	49.2	4.8	502	10	BM335757
5	49.2	4.8	759	10	BF259495
6	48.8	4.7	524	9	AV436476
7	47.6	4.6	463	10	BG946638
8	47.6	4.6	539	10	BM332294
9	47.6	4.6	564	10	BM075274
C 10	47.6	4.6	1101	12	CNS0175V
11	46.6	4.5	497	9	AA388278
12	46.4	4.5	591	10	BI960325
13	46.2	4.5	513	9	AM927362
14	46.2	4.5	534	9	AM331023
15	46.2	4.5	548	9	AM927363
16	46.2	4.5	490	9	AM520296
C 17	46.2	4.5	545	10	BE511221

18	46	4.5	588	9	AM566229
C 19	46	4.5	712	10	BM351809
20	45.8	4.5	501	10	BI273373
21	45.8	4.5	562	10	BF484304
22	45.8	4.5	762	10	BF265677
23	45.6	4.4	526	12	BH219902
24	45.6	4.4	530	12	BH219884
25	45.4	4.4	560	10	BF478389
26	45.4	4.4	676	10	BG905572
27	44.8	4.4	376	10	BE918312
28	44.8	4.4	664	10	BM328060
29	44.6	4.3	342	9	AM564526
30	44.6	4.3	358	10	BG273412
31	44.6	4.3	388	10	BF656946
32	44.6	4.3	412	9	AM564034
33	44.6	4.3	415	10	BE596249
34	44.6	4.3	427	10	BI074677
35	44.6	4.3	468	10	BG240338
36	44.6	4.3	469	10	BE675548
37	44.6	4.3	478	10	BG240581
38	44.6	4.3	493	10	BG946964
39	44.6	4.3	505	10	BE365880
40	44.6	4.3	525	10	BE594368
41	44.6	4.3	535	10	BE360673
42	44.6	4.3	546	9	AM671463
43	44.6	4.3	558	10	BE726101
44	44.6	4.3	598	10	BI716428
45	44.6	4.3	598	10	BE596557

ALIGNMENTS

RESULT 1
CNS0091P/c
LOCUS
DEFINITION
Drosophila melanogaster genome survey sequence TERT3 end of BAC #
BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION
AL053013
VERSION
AL053013.1 GI:4934461
KEYWORDS
SOURCE
ORGANISM
Drosophila melanogaster
fruit fly.
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 925)
Genoscope.
JOURNAL
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammose in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp. The same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
Location/Qualifiers
1..925
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"

BASE COUNT	120 a	61 c	61 g	172 t	511 others
ORIGIN	/clone="BACRI9D16" /note="end : TE73"				

Query Match	5.38;	Score 54.6;	DB 12;	Length 925;
Best Local Similarity	15.88;	Pred. No. 0.099;		
Matches 63;	Conservative 163;	Mismatches 172;	Indels 0;	Gaps 0;

[illegible]

LOCUS	DEFINITION	925 bp DNA	linear	GSS 03-JUN-1999
CNS0091P	Drosophila melanogaster genome survey sequence TET3 end of BAC # BAC119p16 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.			

ACCESSION	AL053013
VERSION	AL053013.1
SYNWORDS	GSS
ORGANISM	fruit fly. Drosophila melanogaster

REFERENCE	AUTHORS	TITLE	JOURNAL
-----------	---------	-------	---------

COMMENT

Submitted (02-JUN-1999) Genoscope - Centre National de Séquençage
BP 191 91006 Evry cedex - FRANCE (E-mail: seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)

determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the *Drosophila melanogaster* genome using these BACs. For further information please see <http://www.fruitfly.org/TheBDGP/Drosophila/melanogasterBAClibrary>

BAC library was prepared by Kazuo Osogawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPc1-98 and was constructed by partial EcoRI digestion of *Drosophila* DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be

```

FEATURES
source      found at http://bacpac.med.buffalo.edu/drosophila_bac.htm
            Location/Qualifiers
            1..925
            /organism="Drosophila melanogaster"

```

BASE COUNT	120 a	61 c	61 g	172 t	511 others
ORIGIN					

Query Match	5.0%;	Score 51.8;	DB 12;	Length 925;
Best Local Similarity	13.7%;	Pred. No. 0.39;		
Matches 54;	Conservative 172;	Mismatches 167;	Indels 0;	Gaps 0;

QY	403	ccctcggagatgcttggcccttaaccttggaggttaaccacggagacacgtccacgccttgaag	462
Db	532	CSMSCKGKCCGATSTBGTSTTTTSTSSGATGKCGSGGGBSCSCSSCSCSSCSCBCC	591
QY	463	aagtcctctgagaaaaagygacctgltgycacagcagtcctctcaaccaacgctcaacagg	522
Db	592	CCCCSSSYCCSSSBSSSMKCSSTBSGCCCKSKVCGTSCSSSSCSCSSSTSSSSTSS	651
QY	523	gagcgccgggccatcgcgfacaccttgggcgtcccgctctgagccagggaaagccagctc	582
Db	652	STSKSSSSGSSSSSSSTTTKSTASGSGSWASGSGSGTGSTSSSSSSSTSTSSSV	711
QY	583	accctggagcgaatacatctacccctggagagacgtccctcaagacatgycgaacaggcgtg	642
Db	712	SSGSKSTBSBSGSBSSSSSSSTSSBSCSTSTSSSSSSSYSSSTSCCTCCCSYSS	771
QY	643	ctctctctcaactggtgtcaaaagcctbaccagagacacaggaalccgcgcccaaccttgagctg	702
Db	772	STSSSSSTSTSWGTSSTSSSSSVGTSSSSDSTSTSCCCCMCTCSTYBMACYSTSCGGS	831
QY	703	ctggtctctcttggtgctcaggaggaaaggtgtatgcccaacaccaaagacgtggccgttgac	762
Db	832	SSSSCKGKGVTTCGCCGCGSSSTNGMBGTSMSKCSSSSBSCCSSSVSSSSSKSSASSSVSSS	891
QY	763	ctgggcctcatccttgctctcctcccaagttgag	795
Db	892	GSAGVSNMSSASKSSSGSVSSGSGSGSGSVS	924

RESULT	3				
BI960118					
LOCUS	BI960118	637 bp	mRNA	linear	EST 22-OCT-2001
DEFINITION	BI960118				
	HVSNSEN0023E18f	Hordeum vulgare	racinis EST	library	HVCDNA0015
	(normal)	Hordeum vulgare	CN4 clone	HVSNSEN0023E18f,	mRNA sequence.

ACCESSION	B1960118	
VERSION	B1960118.1	GI:16311373
KEYWORDS	EST.	
SOURCE	barley.	
ORGANISM	<i>Hordeum vulgare</i>	

REFERENCE

JOURNAL
COMMENT

Wing, R., Close, T.J., Kleinborts, A., Wise, R., Chin, A., Begum, D.,
Frisch, D., Atkins, M., Yu, Y., Henry, D., Palmer, M., Rambo, T., Simmons,
J., Oates, R. and Main, D.
Development of a genetically and physically anchored EST resource
for barley genomics: Morex rachis cDNA library
Unpublished (2001)
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: twing@clemson.edu
Total hg bases = 579

Seq primer: AATTAACCTCACTAAGCG
High quality sequence start: 5
High quality sequence stop: 633.
Location/Qualifiers

FEATURES

Source

1.637
/organism="Hordeum vulgare"
/cultivar="Morex"
/db_xref="taxon:4513"
/clone="HVSME0023E18f"
/clone_lib="Hordeum vulgare rachis EST library HVCNDA0015 (normal)"
/tissue_type="rachis"
/lab_host="TJC121"
/note="Vector: pBluescript SK(-); Site.1: EcoRI; Site.2: XhoI. Plants were grown at Washington State University, Pullman, WA in a greenhouse, the rachises were excised and frozen in liquid nitrogen (Kleinholz lab). In the TJC Close lab at the University of California, Riverside total RNA was prepared, poly(A) was purified, one primary unamplified cDNA library was made, and 1 million pfu were in vivo excised to give pBluescript SK(-) cDNA phagemids (Chin). Phagemids were plated and picked at the Clemson University Genomics Institute (CUGI) (Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA preparations, DNA sequencing and sequence analysis were performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Rambo, Main). The sequence has been trimmed to remove vector sequence and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and sequence analysis see
http://www.genome.clemson.edu/projects/barley. To order this clone see http://www.genome.clemson.edu/orders Also see Close TJ, Wing R, Kleinholz A, Wise R (2001) Genetically and physically anchored EST resources for barley genomics. Barley Genetics Newsletter 31:29-30. (http://wheat.pw.usda.gov/g9pages/dgn/31/cover.html)"

BASE COUNT 125 a 249 c 158 g 105 t
ORIGIN

Query Match 4.9%; Score 50.6; DB 10; Length 637;
Best Local Similarity 46.3%; Pred. No. 0.67;
Matches 167; Conservative 0; Mismatches 194; Indels 0; Gaps 0;

447 cgtccacgctggaagaagctcttgaagaagaagccttggtccacgacgtcctca 506
133 CATCAACCGGAGAACTACAGCTGGGGCTCCGTCATGAGAAAGCCCGGCGCA 192

507 ccaaacgtcaacgagcgccgagccatcgaccccttggcgctgagggcc 566
193 CAGATCGACTTCGGGAGGGCGGCGCTCCGTCGTGAGCGCCCTCGAGAGCA 252

567 agggaaagcagcagctcaacctgaactatcacctctggaagcctgcctaga 626
253 GGCCAAACCAAGGACCTTCCTGCTTCGTGAGCGCGACCAAGCAACTACCTCAA 312

627 catgagcaacgagcggtgtctcccttcaactggttgaagcctaccagagcaacggaatcgcg 686
313 CTACACGACGCGCTCATGAGCTCGTCAAGGTGGCGGCTCTCGCTACGCAACAC 372

687 ccccaacctgagcgtgtgtctctgtgtcaggggaaagggtgtagtcccaacaccaa 746
373 CCTGTGAGAGGGCTCCGTCGTGCTCCCGCGAGCGCCCATGCGCAAGTACATCCGCTA 432

747 gaccgtgcccgtgacctggtgacctatcctgtctctcccgaggtgtagaggttccaact 806
433 CTACGCGGACTTCGTCCTGACCTCAACAAGGCGCTCGCGCACCAAGCGCGTGAGAT 492

807 c 807
493 c 493

RESULT 4
BM335757/c
LOCUS
DEFINITION
502 bp mRNA linear EST 16-JAN-2002
MEST160-F05.T3 ISUM5-RN Zea mays cDNA clone MEST160-F05 3', mRNA

ACCESSION
BM335757
VERSION
BM335757.1 GI:18165918
KEYWORDS
EST.
SOURCE
Zea mays.
Zea mays.
Zea mays.

REFERENCE
AUTHORS
TITLE
Wen.T.J., Qiu.F., Guo.L., Ashlock,D.A and Schnable,P.S.
1 (bases 1 to 502)
Expressed Sequence Tags from B73 Maize: various stages and tissues including seedlings treated with a variety of hormones

JOURNAL
COMMENT
Unpublished (2001)
Contact: Patrick S. Schnable
Schnable Laboratory
Iowa State University
405 Agronomy, Iowa State University, Ames, IA 50011-1010, USA
Tel: 515-294-0975
Fax: 515-294-2299
Email: schnable@iastate.edu

Individual basecall and confidence value were assigned using the phred software.
(<http://depts.washington.edu/ventures/collabtr/direct/index.htm>#rt). Overall sequence quality assessment and vector trimming were conducted using the Lucy software (<http://www.figr.org/softlab/>). Lucy parameters were set to ensure an overall trimmed quality of 97.5% or better without any vector fragments in the chosen high-quality region of each sequence. Low-quality bases between the poly-T and the high-quality region were replaced with N's to serve as spacers.

PCR Primers
FORWARD: primer T7-1 (AA TAC GAC TCA CTA TAG)
BACKWARD: primer T3 (ATT AAC CCT CAC TAA AG)
Seq primer: primer T3 (ATT AAC CCT CAC TAA AG).
Location/Qualifiers

FEATURES

source

1.502
/organism="Zea mays"
/cultivar="B73"
/db_xref="taxon:4577"
/clone="MEST160-F05"
/clone_lib="ISUM5-RN"
/tissue_type="mixed"
/lab_host="DH10B"
/note="Vector: pT73PAC; Site.1: EcoRI; Site.2: NotI. Tissues: Germinated seed and seedlings (1, 2, 8, 11 DAG), Mixed mature tissues (17, 21, 38, 69, 77 DAG), kernels (3, 5, 10, 15, 20, 25, 30, DAP), Adventitious roots (65 DAG), tassels (3-39 cm, 53 and 56 DAG), Immature ears (0.2-3.0 cm, 33, 36, 59 DAG), Husk (73 DAG), Silk, unpollinated first ear, ear shank, etiolated seedlings, callus, Cycloheximide-treated callus, Anaerobic treated seedlings, NAA (a-Naphthalene acetic acid)-treated seedlings, Kinetin-treated seedlings, AACC (1-aminocyclopropane-1-carboxylic acid)-treated seedlings, Brassinolide-treated seedlings, ABA (Absciscic acid)-treated seedlings, GA (Gibberellic acid)-treated seedlings, JA (Jasmonic acid)-treated seedlings. ds-cDNA molecules were generated as follows. First-strand cDNA was prepared from oligo-dT selected mRNA by priming with a NotI oligo-dT primer (5' AACTGGAGAGATTGGCGCGCGAGGAAATTTTATTTT). The resulting DNA:RNA hybrid was treated with RNase H and used as a template for DNA PolI-catalyzed second strand synthesis. After the addition of EcoRI adaptors, the ds-cDNAs were digested with NotI and size-selected. The resulting molecules were directionally cloned into the EcoRI and NotI sites of the pT73PAC vector. The library then went through one round of normalization to COT value

TITLE	Generation of 10,154 expressed sequence tags from a leaky gametophyte of a marine red alga, <i>Porphyra yezoensis</i>
JOURNAL	DNA Res. 7, 223-227 (2000)
MEDLINE	20363100
COMMENT	Contact: Eriika Asamizu The First Laboratory for Plant Gene Research Kazusa DNA Research Institute Yana 1532-3, Kisarazu, Chiba 292-0812, Japan Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/ Location/Qualifiers
FEATURES	
source	1. 524 /organism="Porphyra yezoensis" /strain="TU-1" /db.xref="taxon:2788" /clone="PS002f12_r" /clone_11p="Porphyra yezoensis TU-1" /note="Vector: phluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT	91 a 176 c 196 g 61 t
ORIGIN	
Query Match	4.7%; Score 48; DB 9; Length 524;
Best Local Similarity	45.6%; Pred. NO. 2.3;
Matches 208; Conservative	0; Mismatches 245; Indels 3; Gaps 1;
Oy 284	aagcgttcacgagcagcggtcctcaagaatctcttgaacctctccaaagagatgccgctcca 343
Db 69	AGCGCGCCGAGCGCGCTCGCGCGCGCGCGCTGACTGGAAGACAGACTCGCGCGCTCA 128
Oy 344	ccatccctcaccgcaacccggyagctctcyagaagctggygaagctgctctcatgctcc 403
Db 129	CCCCCGCGAGCGCAAGCGCGCGCGCGCGAGCGGGTGAGAGAGCGGCACTTTGAAAAAG 188
Oy 404	cccttgaaatcttggtcctcaacctggyggtcaccgcgcagaccgttcaagctctgaaga 463
Db 189	CCAAAGAAATGTTTTCACACCGCCCGCGCTGGCGGCGCGAGCGCGCGCGCGCTGACCG 248
Oy 464	aggtccttgagaaaaagggcctgctgctgacccagcagctcccttaccaaaacgctcaacg 523
Db 249	ACTCATCGACACCTTCCCGCCCGCGCGCGCTGCGAGCGGAGCCACGCGCGCTTTGGCGGC 308
Oy 524	agcgcgcgggc---atcgacacccctttggccgtccgctgtgaagccaaaggaagccagc 580
Db 309	TGCTGGGCGCCAAAGTGTGCGCACCGCTCTGGTCCGCGCGCGCGCGCTTCAAGTACGTGGGA 368
Oy 581	tcaacctgagcagctacatctacccctgaggaagacctgacctagatgacatgacacagcg 640
Db 369	TGCTAAAGGGCATCTCCGCGACCGCGCTGCGGACTGTGCGGACGACGCTCAAGAGAC 428
Oy 641	tgtctctcttcaacttgygtcaaggtctacagagcacaagaaatccgcccaccccttgagc 700
Db 429	TGCTGTGCGGTGCAACGAGCTGCGGAATGAAAGAACTCAAGAGAGGGTCTCAAGCGGGA 488
Oy 701	tgtctgtctcctggtgctcaggggaaaaggtatctc 736
Db 489	AGGGGAGAGAAACCGGTGAAGAGAGTCTGC 524
RESULT	7
LOCUS	BG946638/c
DEFINITION	463 bp mRNA linear EST 12-JUN-2000
ACCESSION	949012011.X1 949 - Juvenile leaf and shoot cDNA from Steve Moose
VERSION	BG946638
KEYWORDS	Zea mays cDNA, mRNA sequence.
SOURCE	EST.
ORGANISM	Zea mays. Zea mays. Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoidae; Andropogoneae; Zea. 1 (bases 1 to 463)

FEATURES	source
AUTHORS	Walbot, V.
TITLE	Maize ESTs from various cDNA libraries sequenced at Stanford University
JOURNAL	Unpublished (1999)
COMMENT	Contact: Walbot V Department of Biological Sciences Stanford University 855 California Ave, Palo Alto, CA 94304, USA Tel: 650 723 2227 Fax: 650 723 8221 Email: walbot@stanford.edu Plate: 949012 row: C column: 11. Location/Qualifiers 1. .463 /organism="Zea mays" /cultivar="W64A" /db_xref="taxon:4577" /clone_id="949 - Juvenile leaf and shoot cDNA from Steve Moose" /tissue_type="Immature leaf primordium and vegetative meristem"
BASE COUNT	79 a 132 c 175 g 77 t
ORIGIN	
Query Match	4.6%; Score 47.6; DB 10; Length 463;
Best Local Similarity	48.9%; Pred. No. 2.8;
Matches 128; Conservative	0; Mismatches 134; Indels 0; Gaps 0;
QY	344 ccatccctctacggcaaccgagagctctggaaggtgaggacggttcgtctcatggtcc 403
DB	358 CCACCCCTACGACACCGCCTCCACATCTGGAGAGTGACTACACGCTCCGTCGCGC 299
QY	404 ccttgagatgtgtgcccctcaaccctgggggtgacacggagacggtccacgctggaaga 463
DB	296 GCAAGGAGGTGGCGTCCGGCACCACTGCGGACCCGGCTGGATGCGCGCCAGCGCTCCA 239
QY	464 aggtctcttaaaaaagggcgtgctggtggccacgacgctcttaccacaaacggtcaacgg 523
DB	238 CCGCGCTGGAGATCCCGCCGCAAGGAGCCCTACGACTTCCTCGTCTCCCTCGTCAGGAGACG 179
QY	524 agcgcgcgggcatcggaaccttgggcccgtccgctgagagcccaaggaaagccaggctca 583
DB	178 TCGGGCGGGAGCGGAGACATCGACTCGAGAGCTCAGGTCGGGCTCACCGCTGCACTCCCA 119
QY	584 cccttgagcactacatctacc 605
DB	118 TCGTCGGCAGCACTTCACCATCC 97
RESULT	8
LOCUS	BM332294
DEFINITION	MEST154-H06.73 ISUMS-RN Zea mays cDNA clone MEST154-H06 3', mRNA
ACCESSION	BM332294
VERSION	BM332294.1
KEYWORDS	EST; GI:18162455

/note="Vector: pTR33PAC; Site 1: EcoRI; Site 2: NotI;
 Tissues: Germinated seed and seedlings (1, 2, 8, 11 DAG),
 Mixed mature tissues (17, 21, 38, 69, 77 DAG), kernels
 (3, 5, 10, 15, 20, 25, 30, DAG), Adventitious roots (65 DAG
), Tassel (3-39 cm, 53 and 56 DAG), Immature ear (0.2-3.0
 cm, 53, 56, 59 DAG), Husk (73 DAG), Silk, unpollinated
 first ear, ear shank, etiolated seedlings, callus,
 Cycloheximide-treated callus, Anaerobic treated seedlings
 , NAA (a-Naphthalene acetic acid)-treated seedlings,
 Kinetin-treated seedlings, ACP
 (1-aminocyclopropane-1-carboxylic acid)-treated seedlings,
 Brassinocyclopropane-1-carboxylic acid
 (Brassinolide)-treated seedlings, ABA (Abscisic acid
)-treated seedlings, GA (Gibberellic acid)-treated
 seedlings, JA (Jasmonic acid)-treated seedlings. ds-cDNA
 molecules were generated as follows. First-strand cDNA was
 prepared from oligo-dT selected mRNA by priming with a
 NotI oligo-dT primer (5',
 AACTCGAAGAATTCGCGCGCAGCAATTTTTTTTTTTTTTTT). The
 resulting DNA:RNA hybrid was treated with RNase H and used
 as a template for DNA polymerase II second strand
 synthesis. After the addition of EcoRI adaptors, the
 ds-cDNAs were digested with NotI and size-selected. The
 resulting molecules were directionally cloned into the
 EcoRI and NotI sites of the pTR33PAC vector. The library
 then went through one round of normalization to Cot value
 of 5 based on the methods of Marcello Bento Soares (Genome
 Research 6: 791-806, 1996)."

Query Match	Similarity	4.6%	Score 47.6	DB 10	Length 564
Best Local	Similarity	48.9%	Pred. No. 2.9		
Matches	128	Conservative	0	Mismatches 134	Indels 0
				Gaps	0
OY	344	ccatccctcaagcgaacccgggaagctctggaagaagtgaggacgctgcgtcatcgtatcc	403		
DB	503	CCAACCCCTACGACACACCGCCCTCCCATCTGGAGGTGACATACACGCTCCGGTGGCGG	444		
OY	404	cccttgaaatgtgtgcccctcaaccctgggggtgcaccggcaagacggtccagcttgaaga	463		
DB	443	GCAAGAGAGTGGCGGTCCGGACCATATGCCGAGACCCGGCGTGATGCCCGCACGGCTCCA	384		
OY	464	aggtccttggaagaaaggagcctggtggtgcaccggaacgtcttcacccaagacgylcaaggg	523		
DB	383	CCGCGCTGGAGATCCCCGCAAGGTGCCCTAGCATCTCTCTCTCCCTCCGACAGGAGG	324		
OY	524	agcgccggggccatcggaccccttgggcctgctcggtctgagggcagaaggaagccaggtcca	583		
DB	323	TCGGCGCGAGCTGGGAGACATCGACTACGAGCTCTCAGAGTGCGGCTCACGCTGCACCTCCCA	264		
OY	584	cccttgacgaactacatctacc	605		
DB	263	TCGTCGGCACTTCACCATCC	242		
RESULT	10				
CNS0175X					
LOCUS	CNS0175X	1101 bp	DNA	linear	GSS 26-JUL-1999
DEFINITION	Drosophila melanogaster genome survey sequence SP6 end of BAC				
	BACN37J08 of DrosBAC library from Drosophila melanogaster (fruit				
	fly), genomic survey sequence.				
ACCESSION	AL108460				
VERSION	AL108460.1	GI:5628764			
KEYWORDS	GSS.				
SOURCE	fruit fly.				
ORGANISM	Drosophila melanogaster				
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;				
	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;				
	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.				
REFERENCE	1 (bases 1 to 1101)				
AUTHORS	Genoscope.				
TITLE	Direct Submission				

JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)

COMMENT
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelOBAC11.

FEATURES
 source Location/Qualifiers

 1..1101
 /organism="Drosophila melanogaster"
 /plasmid="pBelOBAC11"
 /db_xref="taxon:7227"
 /clone_id="DrosBAC"
 /clone="BACN37L08"
 /note="end : SP6"

BASE COUNT 254 a 176 c 160 g 152 t 359 others

ORIGIN

Query Match 4.6% ; Score 47; DB 12; Length 1101;
Best Local Similarity 11.8%; Pred. NO. 4.3; Mismatches 102; Indels 0; Gaps 0;

Matches 34; Conservative 153; Mismatches 102; Indels 0; Gaps 0;

Dy 516 caacgaggagcgcggcgccatcgacccttgggcccgtcggctgaagccaaggaaagc 575
 |||:: :::: |||| |:: :: :: :: :: :: :: :: :: ::
Db 785 SASCSASAMCGVSGSSSCSASCSCGVCAGVSASVASASSVMSKVASAVASCAVASGM 844

Dy 576 cagctcaccttgtagcaactatactaccttgtagagaacctgcccttagaatgcca 635
 :|||:: :::: |||| |:: :: :: :: :: :: :: :: :: ::
Db 845 SAGAIVSSCCRSSVSAVAASVVSSSSSSSVASAASASSSSAASAAVAIAAAAA 904

Dy 636 cgagcgttccttccttaactgaagttaaaggccaccaggaaccaaggaattccccaccct 695
 ||||:: :::: |||| |:: :: :: :: :: :: :: :: :: ::
Db 905 VAVSSVSAASVVSSSSCCSSSSASVVASVVAASASVSSSSSVSTISSASVSVSA 964

Dy 696 gaacgtgctgtcctccttggtcacaaggagaaaagggtgatcccaaaccacaagacgtg 755
 |||:: :::: |||| |:: :: :: :: :: :: :: :: :: ::
Db 965 VSMSAVVSSSASSSSSSVSVVVAASAAAASAAAAASSSASAVAASVSSSSSA 1024

Dy 756 cgttacctggcgctcatcctgtgtcctcccgaggtgagcgctcca 804
 |||:: :::: |||| |:: :: :: :: :: :: :: :: :: ::
Db 1025 SSSSSSSASSSSVSSSSCSCGVSVSSSVSAVAASASASASVSVAA 1073

RESULT 11
AA388278 497 bp mrna linear EST 23-APR-1997
LOCUS AA388278

DEFINITION vc24h07.i1 mouse embryo 11 5dpd Mus musculus CNR Clome

IMAGE 175549 K similar to SW-MRLP_HUMAN P33527 MULTIDROME

REFERENCE RESISTANCE-ASSOCIATED PROTEIN 1.; mRNA sequence.
AA388278

VERSION AA388278.1 GI:2041283

KEYWORDS EST.

SOURCE house mouse.
ORGANISM Mus musculus

Musculi
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 497)
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theisinger,B., Wyllie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
The WashU-HMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Db 319 ACGTGGCGCGGAGGACAGCGTCGCCATCAACACACGAGTGGCGGCGTCAAGGTCAGTCCG 378

635 505

850 cygcgccccaccgagcctacccaggaccgaatccggccccaccct 695
||||| |||||

THIS PAGE BLANK (USPTO)